Analysis of Human-System Interaction For Landing Point Redesignation

AE8900 MS Special Problems Report
Space Systems Design Laboratory (SSDL)
Cognitive Engineering Center (CEC)
Guggenheim School of Aerospace Engineering
Georgia Institute of Technology
Atlanta, GA

Author:
Zarrin K. Chua

Advisors:
Dr. Robert D. Braun, Dr. Karen M. Feigh

May 26, 2009

Analysis of Human-System Interaction For Landing Point Redesignation

Zarrin K. Chua
Georgia Institute of Technology, Atlanta, GA 30332

Dr. Robert D. Braun and Dr. Karen M. Feigh
Faculty Advisors, Georgia Institute of Technology, Atlanta, GA 30332

Despite two decades of manned spaceflight development, the recent thrust for increased human exploration places significant demands on current technology. More information is needed in understanding how human control affects mission performance and most importantly, how to design support systems that aid in human-system collaboration. This information on the general human-system relationship is difficult to ascertain due to the limitations of human performance modeling and the breadth of human actions in a particular situation. However, cognitive performance can be modeled in limited, well-defined scenarios of human control and the resulting analysis on these models can provide preliminary information with regard to the human-system relationship. This investigation examines the critical case of lunar Landing Point Redesignation (LPR) as a case study to further knowledge of the human-system relationship and to improve the design of support systems to assist astronauts during this task. To achieve these objectives, both theoretical and experimental practices are used to develop a task execution time model and subsequently inform this model with observations of simulated astronaut behavior. The experimental results have established several major conclusions. First, the method of LPR task execution is not necessarily linear, with tasks performed in parallel or neglected entirely. Second, the time to complete the LPR task and the overall accuracy of the landing site is generally robust to environmental and scenario factors such as number of points of interest, number of identifiable terrain markers, and terrain expectancy. Lastly, the examination of the overall tradespace between the three main criteria of fuel consumption, proximity to points of interest, and safety when comparing human and analogous automated behavior illustrates that humans outperform automation in missions where safety and nearness to points of interest are the main objectives, but perform poorly when fuel is the most critical measure of performance. Improvements to the fidelity of the model can be made by transgressing from a deterministic to probablistic model and incorporating such a model into a six degree-of-freedom trajectory simulator. This paper briefly summarizes recent technological developments for manned spaceflight, reviews previous and current efforts in implementing LPR, examines the experimental setup necessary to test the LPR task modeling, discusses the significance of findings from the experiment, and also comments on the extensibility of the LPR task and experiment results to human Mars spaceflight.

Nomenclature

α	Number of landing sites considered
ϵ	Training parameter
Π	Number of Points of interest
B	Press/Release button operator
F_{LA}	Fuel consumption of Landing Area
H	Number of hazards
M_c	Choosing operator
M_{ij}	Matrix of terrain map color values
M_t	Thinking operator
m	Number of cells containing max(M_{ij})
n	Number of cells containing min(M_{ij})
P	Pointing operator
p	Number of cells containing mode(M_{ij})
R_{LA}	Roughness of Landing Area
$r_{(a,b)}$	Radii of fuel ellipse
S_{LA}	Slope of Landing Area
AFM	Autonomous Flight Manager
ANOVA	Analysis of Variance
ATPL	Air Transport Pilot License
CHPM	Computational Human Performance Model
DOF	Degree of Freedom
DV	Dependent Variable
EDGE	Engineering DOUG Graphics for Exploration

EDL	Entry, Descent, and Landing	NASA	National Aeronautics and Space Administration
FAA	Federal Aviation Administration	PFD	Primary Flight Display
HITL	Human-in-the-Loop	POI	Points of Interest
IFR	Instrument Flight Rated	PPL	Private Pilot Licensed
ITM	Identifiable Terrain Marker	PVC	Polyvinyl Chloride
IV	Independent Variable	RO	Run Order
LED	Light Emitting Diode	SA	Situation Awareness
LEM	Lunar Excursion Model	SAGAT	Situation Awareness Global Assessment Test
LIDAR	Light Detection and Ranging	SO	SAGAT Run Order
LPD	Landing Point Designator	VFDE	Vehicle Footprint Dispersion Error
LPR	Landing Point Redesignation	VFR	Visual Flight Rated
MEDS	Multifunction Electronic Display System		

I. Introduction

Since July 16, 1969, a dozen men have been to the surface of the Moon, eighteen people past low Earth orbit (LEO), and several hundred men and women within Earth's orbit. While these achievements are laudable, the international spaceflight efforts aim to expand the reach and capability of human spaceflight programs. Nations such as China[1] and India[2] are developing manned spaceflight programs, while programs such as the United States National Aeronautics and Space Administration (NASA) plans to send humans back to the Moon[3] and to planets such as Mars.[4] To meet these objectives, new technologies and improved spacecraft system designs are necessary to counter the challenges posed. One challenge with returning to the Moon is precision landing near points of interest (POI) on the Moon's surface. This type of landing is different than the approach used by the Apollo missions, which focused on the task of safely landing a man on the Moon, within the vicinity of a carefully chosen landing area. The landing areas were selected based on the limitations of the Apollo lunar lander, which could only reach areas of relatively flat and smooth terrain on the illuminated side of the Moon. However, the new lunar lander is required to have complete surface accessibility (both illuminated and shadowed regions) and the capability of landing in hazardous terrain. This lander performance is expected to be similar for human Mars exploration - but will be more difficult to achieve due to the limitations in current entry, descent, and landing (EDL) technology.[5]

Despite the fairly unified agreement on why humans should be sent to the Moon and Mars and the technology required to achieve this aim, a greater discord exists on the role of the crew during mission phases such as EDL. Arguments against some form of manned control include increased mission risk, or a desire to fly in fuel- or sensor performance- optimal conditions,[6] or a significant increase in system complexity. Counter-arguments refer to the need for some level of human control - as a backup/supervisory role to highly automated systems,[7] to adapt to unanticipated circumstances and environments,[8] or to make decisions based on real-time data rather than mission-decided logic precoded prior to launch.[9] As Neil Armstrong, celebrated commander of Apollo 11, so colorfully stated, "I [was] absolutely adamant about my God-given right to be wishy-washy about where I was going to land".[10] Regardless of party allegiances, the major concern among both groups remains the same: more information is needed in understanding how human control affects mission performance and, most importantly, how to design support systems that aid in human-system collaboration. This information on the general human-system relationship is difficult to ascertain due to the limitations of human performance modeling and the breadth of human actions in a particular situation. However, cognitive performance can be modeled in limited, well-defined scenarios of human control and the resulting analysis on these models can provide preliminary information with regard to the human-system relationship.

This investigation examines the critical case of lunar Landing Point Redesignation (LPR) as a case study to further knowledge of the human-system relationship and to improve the design of support systems to assist astronauts during this task. To achieve these objectives, both theoretical and experimental methods are used to develop a task execution time model and subsequently inform this model with observations of simulated astronaut behavior. This paper briefly summarizes recent technological developments for manned spaceflight, reviews previous and current efforts in implementing LPR, examines the experimental setup necessary to test the LPR task modeling, discusses the significance of findings from the experiment, and also comments on the extensibility of the LPR task and experiment results to human Mars spaceflight.

II. Background and Literature Review

A. Manned Spaceflight Development for Earth and Mars

The knowledge base of the level of human control and the design of spacecraft systems is dependent on the maturity of the current technology and the resulting application. Many studies for each application (ie, targeted planetary body) have implemented or proposed various technologies to support human control. For example, the Space Shuttle Orbiter, which routinely traverses to LEO and performs an Earth re-entry, received an upgrade in cockpit display technology in the mid 1990s. This new cockpit, the Multifunction Electronic Display System (MEDS), or better known as "glass cockpit", progresses from the standard monochromatic display to an array of colors and enhanced graphics. Two identical sets of six displays (one for each astronaut pilot) present more information in a succinct and efficient manner.[11] The tasks and responsibilities of the astronaut pilots has also been examined. Holland and VanderArk performed a task analysis on the Shuttle entry and landing sequence, documenting the approximate vehicle conditions at the time of each task.[12] While the task analysis assisted in the understanding of the requirements to pilot the Orbiter, this study did not compare the result of neglecting a critical task. Studies such as these are useful for the intended application, but are difficult to extend to other scenarios, due to the differences in flight conditions. This difference is most apparent when comparing Earth and Mars EDL. Mars EDL significantly lags behind Earth re-entry and Moon landing. Thus far, no manned spacecraft has landed on Mars and the required technology to complete such a feat has not been flight validated. Recent studies of manned spacecraft to Mars require significantly heavy payloads (40 - 80 MT), far exceeding the limits of current technology. In comparison, the Mars Science Laboratory, slated for launch in 2011, is about 1 MT and pushing the limits of Viking era landing technology, the current methodology for landing on Mars.[13] Preliminary concepts have been proposed for landing large payloads to Mars, such as NASA Design Reference Missions (DRM) which outline the full Earth launch, Mars landing, Earth return.[4] Other Mars EDL investigations have been focused on parametric studies of various architecture elements. Steinfeldt, et al. compared traditional blunt body and slender body shapes of a variety of entry velocities, masses, and deceleration technologies to determine the amount of payload that could be landed on Mars.[14] Christian, et al. also performed a similar study, but instead focused on traditional Viking and Apollo heritage designs across a range of initial conditions.[15] These studies, along with other human exploration investigations, explore the design space on the method to send humans to Mars, but the actual modeling of human impact is fairly limited - at most, G-loading constraints are included as a factor in EDL trajectory design.

Recent studies of Earth and Mars human EDL applications are largely high-level conceptual investigations. Conversely, studies regarding manned lunar landings are predominantly detailed and focused studies. The current effort to return to the Moon has prompted investigations into the improvement of Apollo-era designs and to incorporate existing elements of space-rated technology. The LPR task has garnered significant attention, due to the criticality of the task, the availability of new technology, and the challenge of the current mission objectives. However, appreciation of the current practice of modeling LPR first requires an understanding of the definition and the history of the task.

B. Landing Point Redesignation

The LPR task, in its most fundamental form, is an opportunity for the astronauts to select a final lunar landing site. The LPR task is expected to occur during the latter portion of the lunar trajectory, during the powered descent phase (PDP). In PDP, the lunar module has just performed a braking maneuver, or reverse thrusting of the main descent engines, and is rapidly decelerating. The vehicle performs a pitch-up maneuver, which places the vehicle in an orientation suitable for LIght Detection and Ranging (LIDAR) sensor operation.[16] This maneuver is expected to occur at approximately 1 km in altitude, at a velocity of 100 m/s (nominal trajectory).[17] Figure 1[16] illustrates this mission sequence.

Soon after vehicle pitch-up, the LIDAR sensor begins to scan the expected landing site. The LPR task begins after the Autonomous Flight Manager (AFM) displays the results of the LIDAR scan. The AFM serves two purposes: 1) processing the raw LIDAR sensor data into a form comprehensible to the crew, and 2) from this sensor data, suggesting alternative landing sites to the *a priori* baseline site. Up until this point, the only source of terrain information available to the crew is a window/synthetic camera view. After AFM processing, the crew has the window/synthetic camera view and the results of the LIDAR scan as sources of terrain information. The crew evaluates the alternative landing sites, finds a site that satisfies the specified criteria (e.g., safety, fuel efficiency, and nearness to point of interest), and designates the final landing site, which concludes the LPR task. This exercise can be analogized to common everyday scenarios, such as determining where to park one's car. A parking spot located near the targeted location is preferable, but avoiding collisions or accidental scrapes by other cars is highly desired. Furthermore, the car is rapidly running out of fuel, so a quick decision is favored to prevent immobility. This analogy is essentially what occurs during lunar

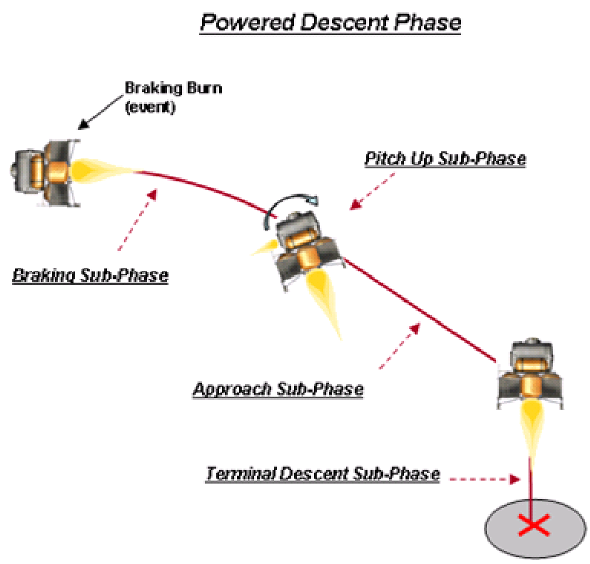


Figure 1: Powered Descent Phase.[16]

landing, but the car is the lunar vehicle, the parking spot is a landing site, and the targeted location is a previously landed asset or a spot of scientific interest. The LPR task generally occurs rather quickly, as the trajectory required to enable LPR is typically not fuel-optimal. Preliminary analyses estimate 30 seconds for LPR task completion.[18] In this short period of time, the astronauts must absorb information from the AFM and window/synthetic camera view, perform tradeoffs of safety, fuel, and nearness to the POI, and select a site. Likewise, the crew must adapt to any unanticipated changes to the terrain.

This current procedure for LPR is different from the landing point selection process utilized during the Apollo program. Apollo-LPR was strictly a manual task and was not a collaboration between an automated system and the onboard crew. The astronauts used a reticle-etched window known as the landing point designator (LPD) to determine the necessary guidance to reach the selected landing site. The site was selected based on the astronaut's perception of the landing terrain, as seen from the window. The pilot would align the cross hairs and view the location of the expected landing site based on readings of the flight computer. The scribe markings are used to retarget the vehicle and land at the new alternative site.

While the LPD system worked for all Apollo missions, this method of landing point redesignation is dependent on the viewing angle, lighting conditions, and familiarity with the lunar terrain. A sound lunar lander design can account for such variations, but there is less control over these parameters. Furthermore, the goals of the Constellation program are to fly in lunar terrain where visibility is poor and hazardous are numerous.[3] The inclusion of the AFM will make the new lander system more robust to environmental factors and increase the lunar lander capability.

C. Current Efforts in Lunar Landing Development

With the shortcomings of the Apollo methods and over four decades of improved technology, recent studies related to manned lunar landing have focused on the development of support systems (ie, displays and landing algorithms) and improved modeling of human performance. There are several key efforts pertinent to this study. The main investigator of landing algorithms is the Autonomous Landing and Hazard Avoidance Technology (ALHAT) team, led by NASA Johnson Space Center. The ALHAT team is developing the AFM and assisting in the design of the associated displays to provide situation awareness and task assistance. The AFM is analogous to a Flight Management System (FMS) - providing guidance, navigation, and control cues,[16] monitoring system health,[19] and interacting with the crew, including prompts for supervisory commands.[20] The ALHAT team has also provided suggestions for cockpit display designs,[21] in particular, for the primary flight displays and horizontal situation indicator. New technologies such as Heads-Up Display (HUDs) are considered for relevancy to the lunar landing situation. Another key proposer of lunar cockpit display designs has been the Altair team. The Altair team, also led by NASA Johnson Space Center, is tasked with the design of the lunar lander. Recent studies such as the work of Prinzel, et al. have capitalized on the maturation of technologies such as Highway in the Sky (HITS).[22]

Several key studies have focused on various elements of LPR. Forest, et al. developed a landing site selection algorithm that, when given terrain data, would highlight key hazards and suggest alternative sites based on the cost function preference of the crew. This study provided an initial reference LPR display, but did not model human interaction with such a system.[23] Needham investigated the impact of varying levels of automation on human performance during LPR, concluding that higher automation allowed for quicker time to complete.[24] In addition, Needham developed a set of icons that would overlay landing site terrain characteristics on a top-down synthetic map. An experiment was also performed to observe the impact of varying levels of automation. However, the subjects used in this experiment were graduate students with little piloting experience and not closely representative of astronauts. Sostaric and Rea modeled the impact of LPR on trajectory design.[6] This impact exists primarily in the need for a window viewing angle and required vehicle divert capability. This study provides initial estimations on the change in metrics (e.g., time to touchdown, trajectory profile) but only for a high-level concept of LPR; no further commentary on specific human tasks is provided. Lastly, Chua, et al. have derived a task model and used this model to examine bottlenecks of LPR.[25] These bottlenecks were addressed by redesigning the LPR display to simplify the information layout and to utilize new symbolism to represent site characteristics. This LPR task model also incorporates expert decision-making theory[26] to account for specialized astronaut behavior.[18] However, this study is based on theory and lacks observations from equatable subjects.

D. Modeling Landing Point Redesignation

This task model (with slight modifications) and the associated reference display proposed by Chua, et al.[18] are the reference works used in this current investigation. Figure 2[18] illustrates the task flow diagram.

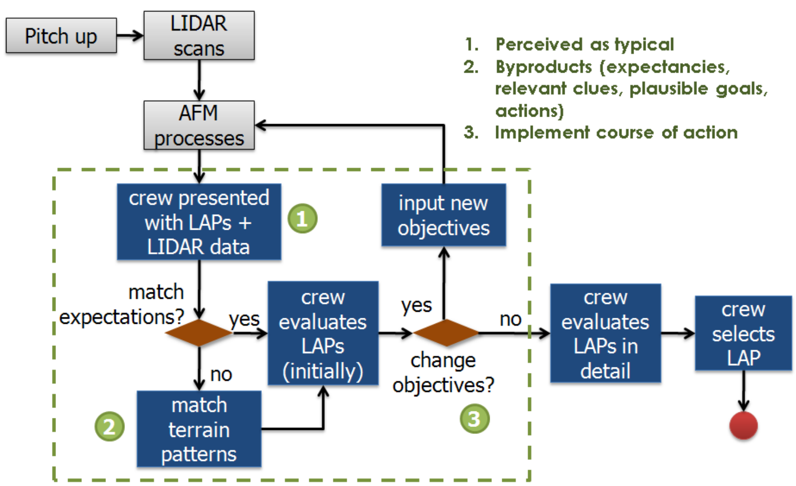


Figure 2: Task Flow Diagram for Landing Point Redesignation. Landing Aimpoints are referred to as LAPs.[18]

This model utilizes the Keystroke Level Model of the Goals, Operator, Methods, Selection Methodology (KLM-GOMS) proposed by Card, et al.[27] The KLM-GOMS methodology decomposes a task into a series of smaller subtasks, until the action can be described with the application of primitive operators (see Table 1) that describe the interaction of a human with a system. All primitive operators are from Card, et al.[27] unless otherwise noted. These operators are different from those used in the original formulation of the task model. Lower operator times are utilized to better account for highly trained astronauts that will be well acquainted with the physical layout of the LPR display and cockpit controls. A summation of all the primitive operators gives an approximate estimation of the time to complete the actual task, or a task execution time prediction model. The model utilized in this study is presented in Equation 1:

$$T(\alpha, \epsilon, \Pi, n, H) = 2.4\epsilon + 4.8\alpha + 2.02n + \frac{6(n+1)(H+\Pi+2)}{5} + 5.72 \tag{1}$$

where α is the number of alternative sites evaluated in detail (including the baseline point); ϵ is the training parameter, where ϵ is 0 if the terrain training matches the actual terrain, 1 if the astronauts are unprepared for the actual terrain; Π is the number of POI, n is the number of objective changes, and H is the number of ITMs[18,25] .

Table 1: Examples of Primitive Operators.

Operator	Symbol	Execution Time, s
Mental activity		
Thinking, perception	M_t	0.62
Choosing[28]	M_c	0.62
Point mouse	P	0.80
Press/Release mouse button	B	0.08

III. Experimental Setup

Although the fundamentals of LPR are deceptively simple, simulation is required to establish an analogous working environment suitable for task modeling and testing. This simulation is composed of two major forms: hardware and software modeling. The driving requirement for both forms is identical: To develop and implement an environment that sufficiently emulates the true scenario experienced by the crew during LPR. In particular, the software simulation is driven to model the LPR reference display, AFM guidance algorithm, and test the experiment hypotheses; the hardware simulation is geared toward exhibiting the realism of next generation lunar landers. The following sections elaborate on the respective hardware and software setup.

A. Hardware development

The hardware for this simulation consists of an external PVC frame and an interior metal-wood-posterboard cockpit structure. The exterior frame is composed of a PVC rectangular prism covered with black cloth. The cloth-covered prism creates an enclosed area that blocks alternative sources of light and nullifies non-experiment audio and visual scenery. This physical enclosure most closely mimics the encapsulated nature of lunar landers and allows for distinctive borders to a controlled local lunar landing environment, drawing users' attention to the interior cockpit structure and subsequent displays. The interior structure begins with a composition of metal frame members and wood paneling to create two shelves for equipment placement. The upper and lower shelves are further elevated to reach the appropriate levels necessary for the experiment. This setup is illustrated in Figure 3. Units are provided in feet.

The upper shelf supports panels of static instrument displays and associated switches and the lower shelf holds three computer monitors and additional, smaller, static panels. The static instrument displays are replications and adaptations of similar panels found on the Apollo Lunar Excursion Module (LEM). The content of these panels are based on critical elements of the LEM that further convey the realism of this mockup.[29] For example, an abort button is included to emphasize the difficulty of lunar landing and provide an alternative should a suitable landing site not be found. These display panels are kept static to reduce distraction from the main task of LPR. Figure 4 shows the primary flight display (PFD) utilized in this experiment. This PFD is in color, in response to current use of the MEDS,[11] or glass cockpit, that is available on the Space Shuttle Orbiter.[30]

The operability of the lunar lander is further emphasized by the presence of two joysticks. These joysticks, which are mounted on either side of the cockpit, do not provide inputs to the software but act as props. The inclusion of joysticks is prompted by the presence of similar actuators on the Apollo LEM. However, the exact location of these joysticks is due to a size constraint (the width of the joystick platform would not allow for sufficient maneuverability within the cockpit) rather than a desire to fully match the Apollo LEM.

The last piece of equipment is an eyetracker. This experiment utilizes a FaceLAB4 eyetracker from Seeing Machines.[31] Due to equipment availability, this eyetracker was not utilized for the first half of the experiment period, but implemented for the later testing sessions. Figure 5[31] illustrates the eyetracker and the associated computing equipment for operation.

With these major components and those described by Tolbert,[29] the hardware aspects of the mock lunar lander are sufficient to emulate the lunar lander environment. The features of the lunar lander are similar to those of the Apollo LEM, but incorporate elements of current human space-rated equipment. Comparisons to the projected interior of Altair, the current lunar lander, confirmed the validity of this experimental setup.[32] Figure 6 illustrates the final form of this mock lunar lander.

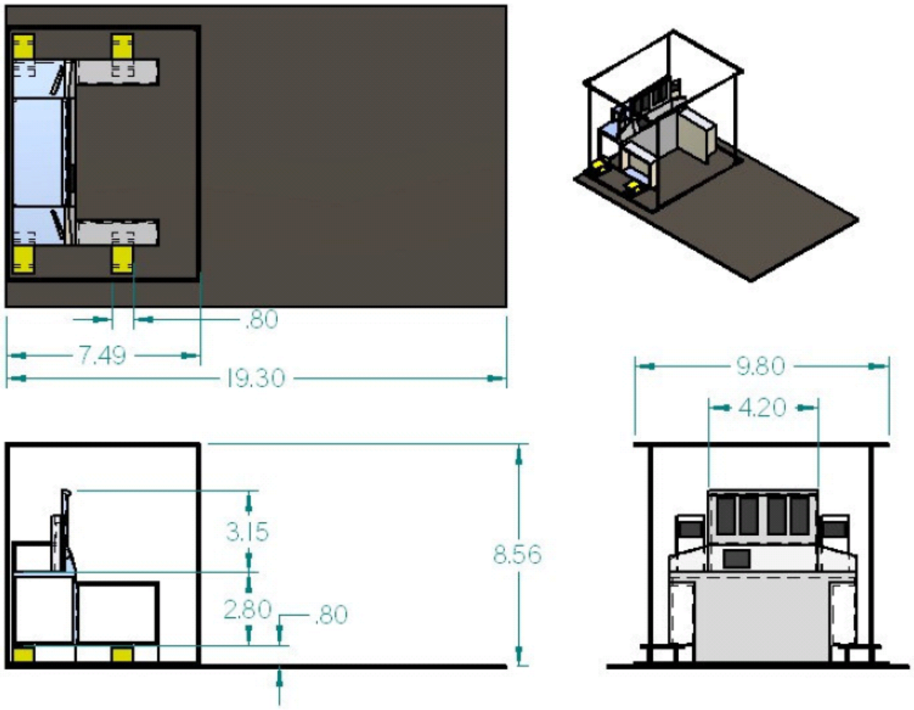


Figure 3: Three View of Mock Lunar Lander Hardware.

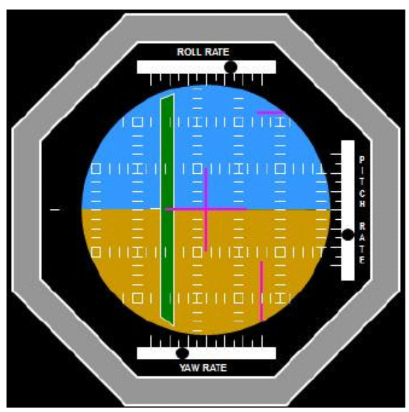


Figure 4: Primary Flight Display.[29]

Figure 5: FaceLAB4 eyetracker from Seeing Machines.[31]

Figure 6: Full Mock Lunar Lander View.

B. Software development

The software development is broken into subcomponents: the from-window perspective, the pseudo-AFM LPR algorithm, and the LPR display.

1. From-Window Perspective

The inclusion of a window on Altair remains a heavily contested issue[6] and is intricately tied to the specific role of the onboard crew. The integration of a window poses non-trivial structural and material property concerns. A window also implies the crew would manually operate the vehicle at some descent and landing stage, thus requiring a need to view the external environment and terrain. A decision on the extent of the role of the onboard crew has not been finalized, but studies such as this and others[19,21,23,24] are providing the data and analysis required to make an informed decision. Regardless of whether Altair chooses to utilize a window view, examination of the Apollo LEM (which included a window that was vital to mission success), the Shuttle Orbiter Cockpit,[12] and relevant aviation applications highlighted a distinct need for a window/synthetic camera view to provide the crew with situation awareness.

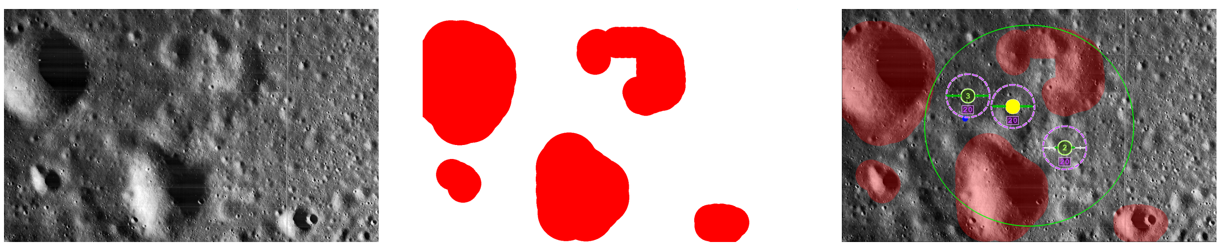
To satisfy this criteria of providing the user with situation awareness, the simulated from-window perspective needed to contain images of the lunar terrain from some altitude and viewing angle. These lunar terrain images must be dynamic, and should display a forward traversing motion. Ideally, these images should also correspond to the inputs of the user, that is, if a divert maneuver is selected, the window view would illustrate a rotation of the local horizon, implying the vehicle has performed a rolling maneuver. However, to implement this level of fidelity is beyond the scope of this investigation. Although this software does exist, as demonstrated by standard spaceflight simulations (Orbiter,[33] EDGE,[34] X-Plane,[35] EagleLander3D[36]), the current state of art still lacks the full breadth of lunar terrain accessibility necessary to uniquely pair window views to final landing site areas. Additional analysis showed that this experiment could achieve similar results with less development time by using a series of coordinated videos. The EagleLander3D Apollo flight simulator software is used to record four approach, landing, and descent scenarios. Each of these scenarios contained from-window perspectives from three stages of flight: vehicle pitchup, vehicle approach, and vehicle landing. The length of video record is fixed to 30, 45, 15 seconds respectively. The vehicle approach video, which played during the tested portion of the LPR task, is played while the user completes the LPR task. If the task is completed sooner than the full video length, then the approach video is stopped and the landing video immediately engages. The initiation of the landing video signals the end of the LPR task.

The from-window perspective is intended for use on the two exterior monitors on the interior cockpit structure. Due to system processing constraints, the monitor on the commander's side of the cockpit is active and the other monitor is kept static. For a more realistic approach, both monitors should be utilized and one monitor should be a reflection of the other (for full symmetry).

2. Pseudo-Autonomous Flight Manager Landing Point Redesignation Algorithm

As previously discussed, the lunar lander is equipped with an AFM that offers alternative landing sites based on an objective function that is set by the crew. Utilizing the alternative site selection algorithm is infeasible; thus, a pseudo-algorithm is necessary to emulate the major aspects of the algorithm, sufficient to provide logical answers without requiring heavy computation. The pseudo-algorithm used in this experiment is based on the AFM algorithm explained by Forest, et al.[23] The algorithm used in this study takes an input package of a landing area (typically satellite photography of the Moon), scans this map for alternative landing sites, and outputs five map images corresponding to each of the following objective functions:[18] safety, fuel efficiency, proximity to points of interest, balanced (equal weighting of the first three), and *a priori*. These functions are discussed in further detail in Section III.B.3.

The input package consists of several image files containing information regarding the specific landing scenario. This package must contain a high quality image of the landing area (this experiment used portable network graphic images of size 1260 × 783 pixels, and is limited by the size of the touchscreen); a hazard map containing just red areas corresponding to major landing hazards; a simplified hazard map with multi-point outline of the hazardous areas; map(s) of the point(s) of interest, designated by a blue circle; and alternative landing site symbols, which include markings for the landing footprint, landing footprint diameter, vehicle cross section, and AFM ranking (1, 2, or 3). Figure 7 illustrates these input files.



(a) Raw Lunar Image. (b) Hazard Identification. (c) Final Image.

Figure 7: Input Files to the Pseudo-Algorithm.

This input package is read and the landing area is converted to a matrix, with each cell containing a value from 0 to 255, greyscale. The pseudo-algorithm treats this matrix as a LIDAR sensor scan, with each cell location corresponding to a geographical position and cell magnitude relating to an altitude. The LIDAR measurement error is not modeled; the resulting image is an idealized LIDAR scan result. The algorithm reads all cell locations that are not marked within the hazardous area and takes a Euclidean vector difference between each non-hazardous point and the point(s) of interest. This vector difference does not take into account hazards between the particular location and the point of interest; this information is expected to be analyzed by the crew. The pseudo-algorithm continues to examine each non-hazardous point by inspecting the landing footprint for safety and fuel consumption requirements. Slope, roughness, and fuel consumption are calculated by Equations 2, 3, 4 respectively.

$$S_{LA} = \max(M_{ij}) - \min(M_{ij}) \qquad i \in [i_0, i_f], j \in [j_0, j_f] \tag{2}$$

$$\begin{aligned} n =& \quad \text{number of cells containing } \min(M_{ij}) \\ m =& \quad \text{number of cells containing } \max(M_{ij}) \\ p =& \quad \text{number of cells containing } \text{mode}(M_{ij}) \\ R_{LA} =& \quad \frac{m\max(M_{ij}) + n\min(M_{ij}) + p\,\text{mode}(M_{ij})}{mnp} \qquad i \in [i_0, i_f], j \in [j_0, j_f] \end{aligned} \tag{3}$$

$$F_{LA} = \frac{\sum_{i=i_0}^{i_f} \sum_{j=j_0}^{j_f} \frac{(i-783)}{r_a^2} + \frac{(j-1260)}{r_b^2}}{(i_f - i_0)(j_f - j_0)} \qquad i \in [i_0, i_f], j \in [j_0, j_f] \tag{4}$$

where the 0 and f subscripts denote the beginning and ending cell component of the expected landing area, r_a and r_b are the radii of the fuel ellipse and M is the matrix of image values. Once these terrain characteristics are computed, the pseudo-algorithm sorts based on the weighting distribution provided for each of the five objective criteria. Logic is included in the algorithm to ensure that unique sites are recommended - no landing site overlaps another. The output map image for each objective function contains hazardous area highlights, the point(s) of interest, the baseline point and three alternative sites, and symbols for the relative goodness of slope and roughness of the expected landing area. These maps are generated prior to the experiment and the map display corresponds to the actions of the user, simulating a real-time calculation of alternative landing sites, without the computational cost or increased risk of system failure.

This pseudo-algorithm also computes the quality of user-selected landing sites based on the criteria presented in each of these maps. This formula, listed in Equation 5, is based on the concept of "the perfect is the enemy of good" (Voltaire). Although the mission wishes to place the lander in an area free of major hazards and of preferable terrain characteristics (flat and level terrain), there exists a region where a significant tradeoff occurs between safety and proximity to points of interest. This region is defined as "adequate" (not violating any tolerances) and to aim for a safer site while drawing farther from the point of interest places the lander at a mission disadvantage. This region of adequacy must be factored into the performance formula, otherwise the resulting ranking is not representative of true lunar landing. Each quality metric (safety - slope, roughness, distance from hazards; fuel consumption; proximity to point of interest) has a region of adequacy.

$$\begin{aligned} P_{\text{score}} &= w_f F_{LA} + D_{POI} + (S_{LA} + R_{LA} + D_h)/3 \\ w_f &= -\frac{\text{Time to Decision}}{\text{Total Time for LPR}} \qquad w_f \in [0,1] \end{aligned} \tag{5}$$

where D_{POI} is the distance from the POI and D_h is the distance from hazards. The element of time is also introduced in this performance formula. Under ideal conditions, the crew would require little to no time to complete LPR as the time to complete the LPR task is tied to other trajectory parameters, such as fuel consumed. The vehicle is assumed to be in a trimmed position and autonomously flown during LPR, to allow the crew to fully concentrate on the task on hand. Trimming the vehicle may require flying in a non-fuel efficient manner, which could ultimately lead to major design reconsiderations such as accommodating a reduction in the payload mass fraction or change in launch vehicle to accommodate lander volume and mass growth. Conversely, restrictions on time to complete LPR may lead to poor decisions, which could lead to mission aborts or increased risk during the final stages of landing. Therefore, the best strategy in performing the LPR task is to select an adequate landing site in the least amount of time possible. Preserving fuel reserves may allow greater surface accessibility or reaching sites on the exterior of the fuel contour. This fuel weighting parameter, w_f is based on the fraction of time-to-decision to allowable time. Faster decisions decrease the importance of the fuel consumption requirement of the selected landing site, whereas slower decisions increase the importance of landing in a fuel-efficient landing site.

3. Landing Point Redesignation Display

The landing point redesignation display, as seen in Figure 8 contains information relevant to the LPR task, the mission status, and methods of communicating to the AFM. This display has a north, center, southeast, and southwest containers (with respect to JAVA Swing terminology). The north container contains three clocks/timers: a mission clock (far left), a redesignation timer (center), and a touchdown timer (right). The southwest container is composed of five buttons. The first four buttons (baseline in yellow, site 1, site 2, site 3) are used to select a final landing site. The button on the far right (Arm) emulates the two-button press method to avoid accidental engagement of a final landing site. To designate a final landing site, the user must press one of the site buttons and then arm. Pressing a site button causes the grey circle to turn yellow (akin to an LED light) and pressing the arm button turns the circle green. The southeast container holds the hot keys, or objective function operator actuators. These hot keys contain predesignated safety and proximity to point of interest tolerances and weighting distributions between safety, fuel efficiency, and proximity to point of interest. These buttons are as defined:[18]

- **Safety**: The safest landing sites (farthest from hazards, conservative tolerances on slope and roughness). Fuel efficiency and nearness to POI are held equal. Example of weight distribution (on a 100 point scale): 90/5/5 or 80/10/10
- **Fuel efficiency**: Most fuel efficient sites (typically center and forward, aft of the LIDAR scanned landing area). Safety and nearness to POI are held equal. Example of weight distribution: 5/90/5, 10/80/10
- **Proximity to points of interest**: Nearest to POI. This objective could be interpreted in different ways, if there are multiple POIs presented: the AFM could find aim points nearest to all POIs, or the AFM could find the closest aim points to at least one POI. This research assumes the later interpretation. Tolerances on slope and roughness are less stringent and safety and fuel efficiency are held equal. Example of weight distribution: 5/5/90, 10/10/80

- **Balanced**: Equal, or balanced weight distribution between safety, fuel efficiency, and nearness to POI. Weight distribution: 33/33/33
- **A Priori**: This distribution is based on mission planning projections of the objectives deemed to be most critical during LPR. This distribution does not include any real-time data. The baseline aim point is based on this weight distribution.

This hot key concept allows for a greater diversity in objective functions without requiring the operator to manually set tolerances and distribution weights. Lastly, the center container consists of a map of the expected landing area and essential information. Information such as alternative landing sites, fuel consumption, slope and roughness, and landing footprint are overlaid on the map to ease crew workload. As noted by Chua and Major,[25] reference displays that are organized in a two panel scheme tend to result in a task bottleneck due to information cluttering and inefficient landing site evaluation. Thus, relative information are placed directly on the map for 1:1 comparison and to minimize eye movement.[37] Table 2[18] lists the major symbols used in this display and their representation.

C. Experimental Procedure

There were three major objectives of the LPR experiment:

1. To inform the current task model by observing the strategies used by experienced pilots.
2. To compare true performance to the predicted task execution times.
3. To determine the efficacy of the current lunar landing redesignation display design.

Pilots were enlisted to participate in the experiment and perform the LPR task in controlled conditions. These pilots were all Private Pilot Licensed (PPL) and had at least 80 hours of flying experience. There were no limitations or restrictions on age, gender, or height. This participant criteria was established to ensure analogous astronaut behavior, without sacrifice to the number of sample data points.

Twenty pilots participated in this experiment, representing a wide variety of flight experience and pilot training. The pilots have flown single- and multi-engine aircraft both for personal and commercial uses. Figure 9 illustrates the distribution of pilot certification and a histogram of the hours of flight experience.

The mean was 277 hours of flying, with a standard deviation of 307 hrs. There were no military pilots participating in this experiment, and only one pilot had experience flying helicopters. The majority of the participants were college students, less than 30 years old. The pilots, unknown to them, were randomly separated into two groups based on ITM frequency. There were eight participants in the (2,4) ITM group and 12 participants in the (1,3) ITM group. Figure 10 illustrates the distribution of pilot certification and flight experience across the two groups.

The mean number of flight experience was 181 hrs in the (2,4) group ($\sigma_{(2,4)}$ = 140.5 hrs) and 340 hrs in the (1,3) group ($\sigma_{(1,3)}$ = 373.1 hrs). The discrepancy in flight hours and inequality of VFR/IFR pilots was not determined until after the experiment, and as such, flight hours and pilot certification are included as covariates in the data analysis. Likewise, the experiment setup was intended to have an equal distribution of pilots for both groups, but there were a few instances where the pilot could not be sufficiently trained to complete the LPR task. There were also instances where the method of data collection was flawed. The data for those pilots are not included in the experimental analysis.

Each testing session lasted two hours. The pilot participation time was broken into three segments: initial briefing/practice, testing, debriefing. The initial briefing provided an introduction to the experiment, the LPR task, and the simulator. The pilots were given approximately 45 minutes of guided practice on the LPR task, in the simulator, where they received feedback on their performance. This practice time allowed the pilots to become comfortable with the simulator and also to formulate strategies. The testing session took about 15 minutes. Each participant completed eight trials in a variety of pre-selected landing scenarios. The scenarios were based on science missions (usually involving sample return, establishing lunar laboratories, or investigating a scientific phenomenon). Every opening training screen included two images from various satellite perspectives of the expected landing area. The training screen is presented in Figure 11. A list of all scenarios used is provided in Appendix VIII. The pilots were also informed that the lander was equipped with a rover, but the capabilities of the rover were limited.

Half of these trials were interrupted for situation awareness testing, using the SAGAT technique.[38] The scenarios were developed based on combinations of the three independent variables:

- *Points of Interest*: *Two-level variable: (1,2)* Landing sites that drive the purpose of the mission. These are either scientifically interesting landing sites or the location of assets.

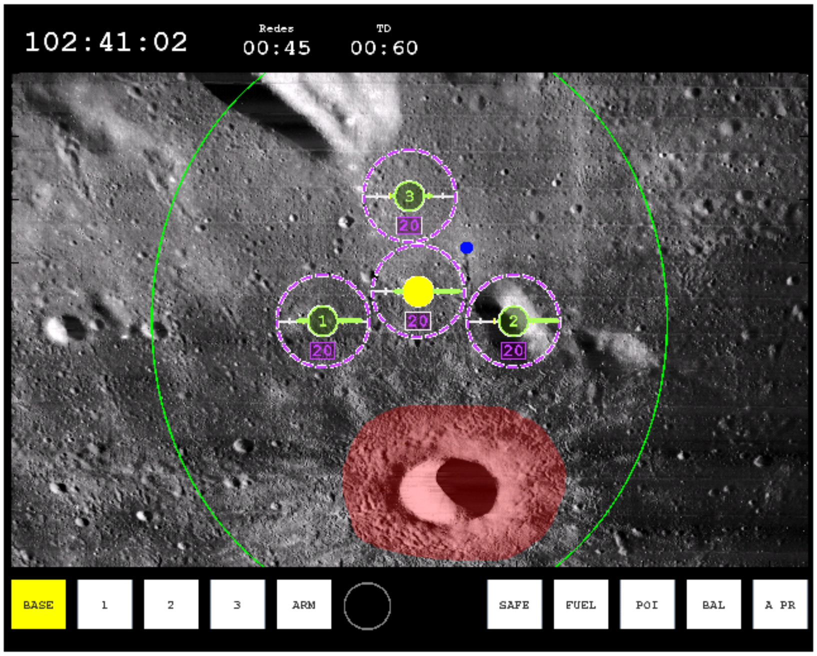
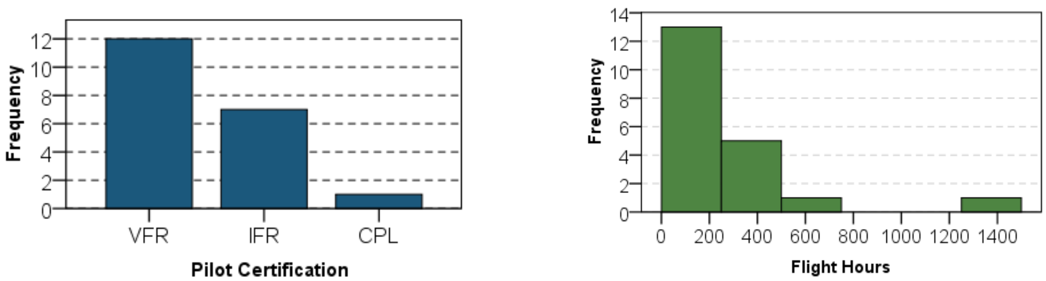


Figure 8: Landing Point Redesignation Display.

Table 2: List of Major Symbols for the LPR Display.

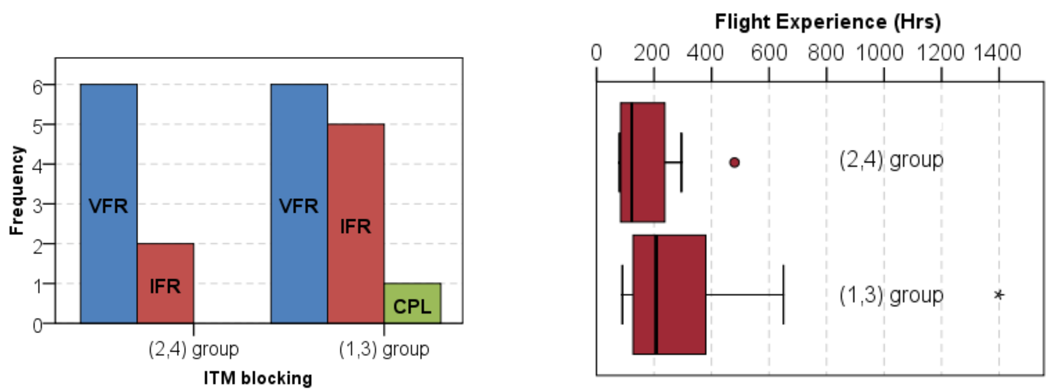
Symbol	Description
Green ellipse	Fuel Contour. The Astronaut Crew Office reported a fuel contour as critical information to execute the LPR task. This display utilizes a green ellipse superimposed on the photo of the landing area to divide the map into reachable and non-reachable fractions. All alternative landing sites are located within this ellipse. This ellipse also represents the relative fuel cost for each landing site. Landing aim points located closer to the center and along the major axis of this ellipse required less fuel than aim points located on the on the fringe and minor axis.
Dashed purple circle	Vehicle Footprint Dispersion Error (VFDE). The VFDE is represented by a dashed purple circle proportional to the area encapsulated on the map. The diameter of the vehicle footprint plus errors is listed in a box in the lower half of this purple circle.
Green circle	Vehicle cross-sectional area. The vehicle cross-sectional area is represented by a green circle located in the center of the VFDE circle. The size of this cross-section is equivalent to the area on the map. Superimposing this information on the landing map also allows the operator to quickly determine the relative distance from hazards and the POIs.
Left and right green arrows	Terrain characteristics. The terrain characteristics of each LAP are represented directly on the map. A modification of the four axis LAP information representation developed by Needham in 2008 is used for this study. Two of the four axes, slope and roughness margins, are utilized. The hazard and fuel margin axes are represented by other symbols. The slope and roughness margin information is displayed in the same manner prescribed by Needham.[24] Three marks along the axes are used to represent dangerous terrain characteristic (defined as, at the threshold), tolerable, and desired (far from the threshold). The arrows are desired to be as long as possible, hence representing a safe LAP.
Blue circle	Points of interest. The points of interest are represented in blue and proportional to the size of this area. Circles and other geometric shapes can be used to represent lunar assets, or scientific spots of interest.
Numbered circles	The numbered circles represent the goodness ranking of the AFM. This numbering also serves the dual purpose of identifying the alternative landing sites while providing the only means of communicating ranked AFM recommendations.



(a) Distribution of Pilot Certification

(b) Distribution of Flight Experience

Figure 9: Demographics of Partipants.



(a) Distribution of Pilot Certification

(b) Distribution of Flight Experience

Figure 10: Demographics of Partipants within the ITM groups.

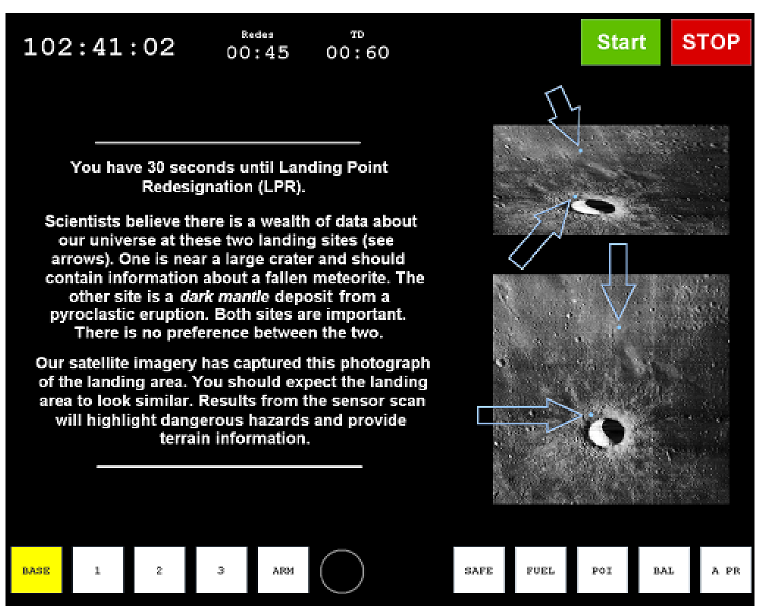


Figure 11: Example of Training Screen Presented Prior to Every Scenario.

- *Terrain Expectancy*: *Two-level variable: (matching, unexpected)* Notes whether the actual terrain matches expectations (i.e., pre-mission training). If the actual terrain matches the training terrain, then the astronauts have better situation awareness than if the terrains do not match. If the terrains do not match, we expect the astronauts to search for Identifiable Terrain Markers (ITMs) through their window, to gain situation awareness. The lunar maps were modified to mimic the variable shading of the terrain. An example of this expectancy is illustrated in Figure 12. The images to the left were presented during the opening training screen and the image to the right represents the actual terrain.

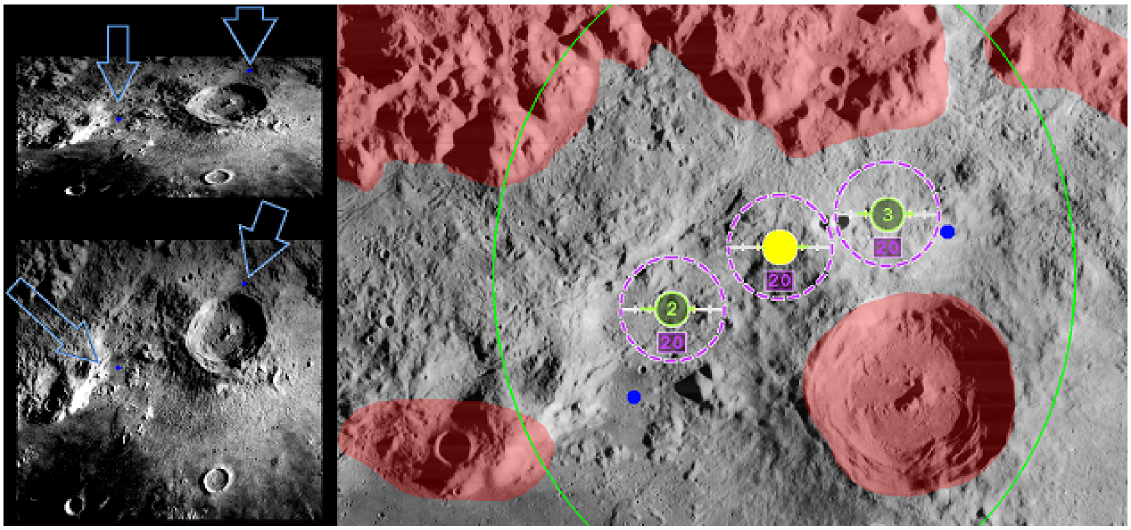


Figure 12: Terrain Expectancy.

- *Identifiable Terrain Markers (ITM)*: *Four-level variable: (1,2,3,4)* Hazards or hazard clusters (within a landing footprint of each other). This variable is also blocked, with participants viewing either 1, 3 or 2, 4 hazard maps.

The full range of scenarios is listed in Table 3.

Table 3: Design of Experiments: $2 \times 2 \times 4$

Scenario	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
POI	1	1	1	1	2	2	2	2	1	1	1	1	2	2	2	2
Expectancy	M	M	M	M	M	M	M	M	U	U	U	U	U	U	U	U
ITM	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4

As mentioned previously, the blocked variable in this experiment is hazards. Thus, participants only experienced half of the scenarios. This sorting and exposure is random. The order of the runs is also random within and between participants. Since several lunar surface maps were reused for different scenarios, the stringent condition for scenario run order was that a map also used for the unexpected scenario could not be shown to the pilot prior to experiencing that particular scenario. Thus, scenarios 10-16 typically occurred prior to their counterparts in scenarios 1-9. This distribution of scenario run order reduced any potential bias in run order (within and between subjects).

This experiment collected several dependent measures. These measures are:

- *Performance*. The participants are given the option to examine 15 alternative landing options, which includes a nominal landing site. The participant can choose to evaluate these options by pressing a hot key to change an objectives (Section III.B.3). These landing sites are ranked relative to each other. A final algorithm computes the performance score, based on the time needed to make a decision and the final landing site selected.

- *Sequence of Events*. The sequence of events is assessed by keeping track of participant actions. Each mission event (such as LIDAR data availability) along with each display interaction is timestamped and recorded. This record is made by the display simulator software. The first and last display interaction is recorded. The sequence of events assists in understanding the strategy, and record the actions completed by the participant.

- *Time associated with each display interaction*. The duration of interaction touch is also recorded. This information aids in primitive operator estimation.
- *Eye fixations*. The eyetracker notes the fixation of the eyes on the display and the window. This data assists in the determine the effectiveness of the display in terms of types and layout of the information.
- *Situation Awareness (SA)*. Half of these runs employ Situation Awareness Global Assessment Testing (SAGAT). The simulation is frozen once and prompts the user to answer questions regarding their current task and working knowledge. The questions are superimposed on the LPR display, which cannot be referred to by the pilots during the freeze. The correctness of the answers to these questions determines the current level of the participant's SA. This test is composed of a variety of questions. There are three levels of SA, all of which are utilized in this experiment: Level 1 - Perception; Level 2 - Comprehension; Level 3 - Projection. A list of the SAGAT questions is listed in Table 4.
- *Display effectiveness*. The display effectiveness is based on the overall performance score (as computed by the algorithm). The participant is also presented with a post-experiment questionnaire, where they are asked to rank the display in terms of quality and are also prompted to provide insights on display elements that worked or failed. The participants are also asked to rate the display based on the Modified Cooper-Harper Evaluation for UAV operator displays.[39]
- *Strategy*. The task model assumes a nominal strategy that is independent on which buttons are pressed, or the strategy in selecting a landing site. Thus, this metric is of particular importance to shed insight on the strategies employed by participants. The participant is exposed to carefully designed scenarios that incorporate different levels of the independent experiment variables. A "talk-aloud" post-experiment briefing method is used for at least one run, to allow the participant to recount his or her specific methodology.
- *Window Usage*. Eyetracker data and post-experiment surveys are employed to determine the window usage.
- *Choice of landing quality metric*. Another interesting metric is the participant-derived definition of landing quality metrics. This choice of landing quality metric is collected by recording the number of hot key presses and from explicit recall during the post-experiment briefing regarding participant strategy.

Additional information regarding the experiment procedure, such as the pilot pre-briefing and debriefing materials, are included in Appendix VIII.

IV. Analysis of Experiment Results

This section discusses the results of the experiment, including the global participant performance and the results of the data analysis.

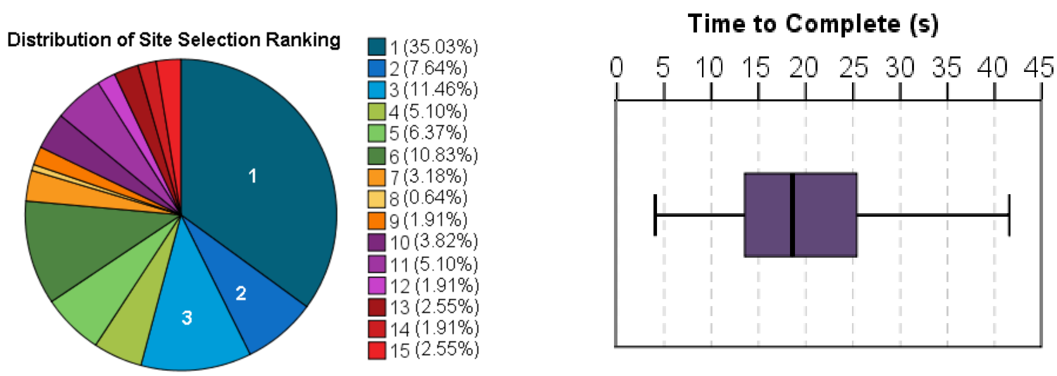
A. Overall Results

Nineteen participants completed all eight runs and the twentieth participant completed five of the eight runs (due to unforeseen simulation failure), for a total of 157 cases across the full design of experiments. The pilots completed the task within the 45 seconds allotted; no pilot had to abort a run. On average, the LPR task was completed in 20.39 seconds ($\sigma = 9.05 seconds$), with the fastest run taking 4.08 seconds and the slowest requiring 41.53 seconds. Overall, the pilots were able to make excellent decisions as to where to land. In 54% of the runs, the pilot chose one of the top three sites, whereas 7% of the runs resulted in the pilot choosing the relatively "worst" sites. While these sites are still feasible landing locations, better sites were available. Figure 13 illustrates the distribution of these site selection rankings and the time to complete the LPR task. Figure 14 ilustrates the same information, but with regard to each pilot certification. The CPL certification is misleading, as there is one only pilot in that category, eight runs total. The VFR pilots succeeded in choosing one of the top three sites in 56% of all their run attempts and the IFR pilots achieved the same level of performance in 51% of their respective runs. The mean times to complete for VFR pilots and IFR pilots were 21.13 (σ_{VFR} = 9.52 s, [5.82, 41.53]) and 18.38 (σ_{IFR} = 18.38 s, [4.08, 36.88]) seconds, respectively.

The time to complete and site selection ranking were compared to determine if there was a significant correlation. Figure 15 illustrates scatterplot of time to complete vs. site selection ranking and the corresponding box plot. Because this data is not normally distributed, a non-parametric test was used to determine if there was a correlation. There is a marginally significant relationship between the two variables, $\tau = 0.085, p = 0.069$. This positive correlation implies

Table 4: Questions for Situation Awareness Testing

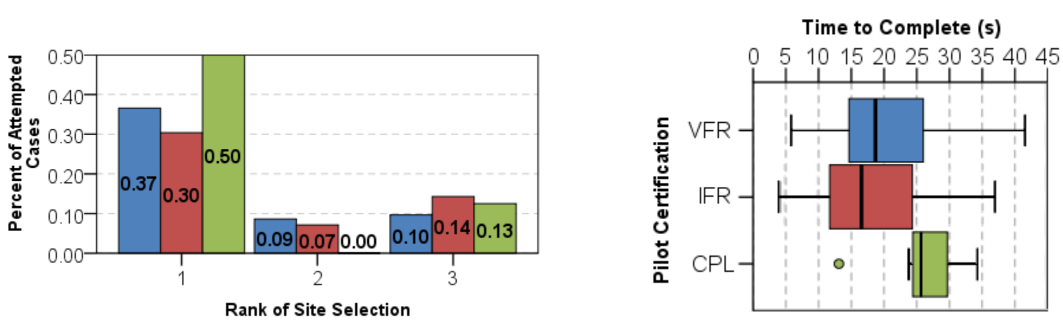
Level	SA #	Question	Choices	Text/Visual?
1	1	How many hazards are on the screen?	1, 2, 3, 4	Text
1	2	How much time is remaining until TOUCHDOWN?	60-51 sec, 50-41 sec, 40-31 sec, 30-21 sec, 20-11 sec	Text
2	3	What percentage of the map is hazard-free?	25%, 33%, 50%, 66%, 75%	Text
2	4	Which landing site, of the ones on the screen, requires the least amount of fuel to achieve?	BASE, #1, #2, #3	Text
1	5	How many landing sites are displayed right now on the screen (baseline plus Alternatives)?	1 site, 2 sites, 3 sites, 4 sites	Text
1	6	How many points of interest are on the screen?	1 point, 2 points	Text
1	7	What is the Landing Footprint Diameter?	5 m, 10 m, 15 m, 20 m, 25 m	Text
2	8	Which landing site, of the ones displayed right now on the screen, is located at the most level spot (best slope)?	BASE, #1, #2, #3	Text
1	9	Where is the baseline site located, within the fuel contour?	More Left, More Right, More Top, More Bottom	Text
3	10	If you were to click on the SAFEST hot key, where do you expect the #1 ranked landing aimpoint to be?	Site A, Site B, Site C, Site D, Site E, Site F	Visual
2	11	Which landing site, of the ones displayed right now on the screen, is located at the smoothest terrain (best roughness margin)?	BASE, #1, #2, #3	Text
2	12	Roughly how close (in terms of landing footprints) is the closest hazard to the points of interest?	1 footprint, 2 footprints, 3 footprints, 4 footprints	Text
3	13	If you were to click on the FUEL EFFICIENCY hot key, where do you expect the #1 ranked landing aimpoint to be?	Site A, Site B, Site C, Site D, Site E, Site F	Visual
2	14	Which landing site, of the ones displayed right now on the screen, is farthest away from hazards?	BASE, #1, #2, #3	Text
1	15	Does the terrain match your training?	Yes, No	Text
3	16	If you were to click on the NEAREST TO POI hot key, where do you expect the #1 ranked landing aimpoint to be?	Site A, Site B, Site C, Site D, Site E, Site F	Visual
2	17	Which landing site, of the ones displayed right now on the screen, is closest to a point of interest?	BASE, #1, #2, #3	Text
1	18	How much time is remaining for REDESIGNATION?	60-51 sec, 50-41 sec, 40-31 sec, 30-21 sec, 20-11 sec	Text



(a) Site Selection Rankings

(b) Time to Complete

Figure 13: Distribution of Performance.

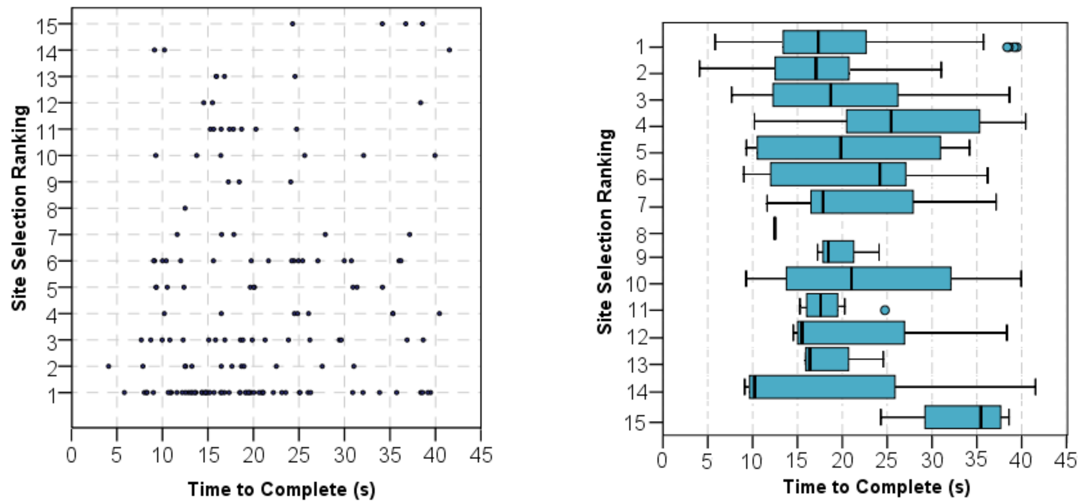


(a) Site Selection Rankings

(b) Time to Complete

Figure 14: Distribution of Performance Per Pilot Certification Category.

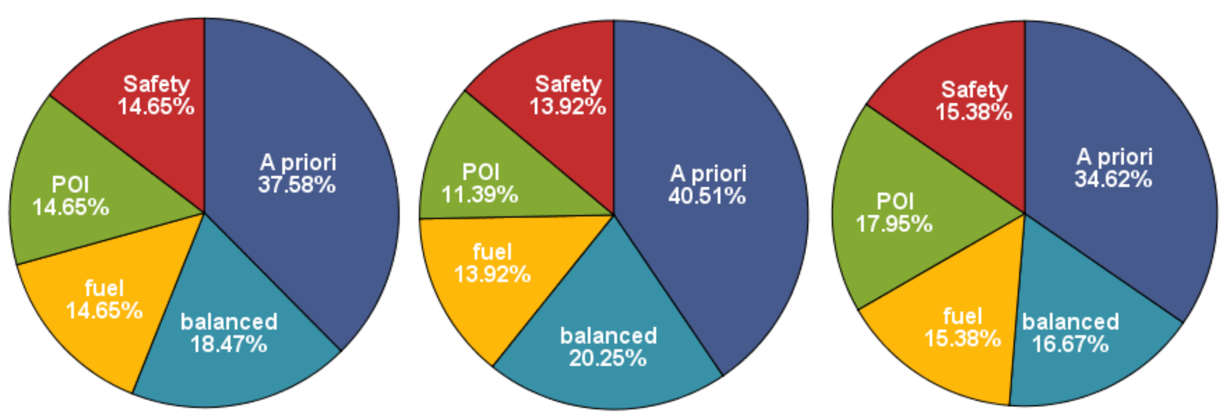
that the longer a pilot takes to make a decision, the worse the overall performance will be. This relationship is intuitive, as the fuel penalty increases with time. However, there are most likely other factors affecting the site selection ranking other than the time to complete.



(a) (b)

Figure 15: Correlation between time to complete and Site Selection Ranking.

In general, the pilots preferred an objective function other than *a priori*. Figure 16 illustrates the percentage of cases where the pilot selected a final landing site from that particular objective function. Selection of the baseline site was included under the *a priori* category, as this site is selected prior to mission launch. However, the baseline point is available as a landing site option regardless of the engaged objective function. The pilots were recorded as selecting the baseline point or one of the *a priori* sites more often than any other selection criteria in unexpected terrain. This observation is not surprising, implying that pilots were more inclined to select a relatively-"known" site in an unknown environment. Predictably, the POI objective function was utilized the least, in only 11.4% of the unexpected terrain cases. This low percentage suggests the pilots were more concerned about their safety or the limits of the landing vehicle than landing near the POI. Conversely, the opposite effect for POI was experienced in expected terrain. Although the *a priori* objective function was utilized most of all criteria, in about 18% of all cases, the pilots seemed to be more inclined to find sites closer to the POI.



(a) Overall (b) Unexpected Terrain (c) Expected Terrain

Figure 16: Preference in Objective Function.

Most pilots found the display useful in completing the LPR task. On a scale from 1-10 (best to worst) on the Modified Cooper-Harper Display Evaluation, pilots generally ranked the display at a score of 2 (*"good with negligible* *deficiencies"*) or 3 (*"Minor but Tolerable Deficiencies"*) (Figure 17). Frequent complaints with the display concerned the representation of information, usually in the category of color or symbol size. Section A discusses in more detail the specific concerns and suggestions for improvement.

The pilots were also given the option to evaluate the window display. This window display, as previously mentioned, contained pre-programmed videos of simulated vehicle pitch-up, approach, and landing sequences. There was no correlation between the maps used for the LPR task and these videos. Most pilots found the window view to be "partly useful", with over half the pilots agreeing with the corresponding statements (Figure 18). This result is contrary to the initial belief that due to the lack of pertinent information available from the window, the pilots would be inclined to rank the window view as "not useful". Pilots' responses to the window view were varied. Several pilots commented on the repetition of the video, and a few believed one video was used for all of their scenario runs. At least one pilot believed there was a correlation between the window and the training maps used for the LPR task, asking out loud if the mountain ridge in one pitch-up maneuver video was the edge of the crater in the satellite footage. In general, most pilots spent the first few runs dividing their attention between the training screen and the window. In the later runs, the window was either glanced at sparingly or ignored entirely. The window was rarely consulted during the actual LPR task. The window view was also where the pilot received notification of LIDAR scan initiation. Some pilots used this information to prepare for the task, settling into different body positions and turning their head toward the LPR display. A few pilots even placed their hand closer to the LPR display, either in the general vicinity of the hot keys or over one key in particular.

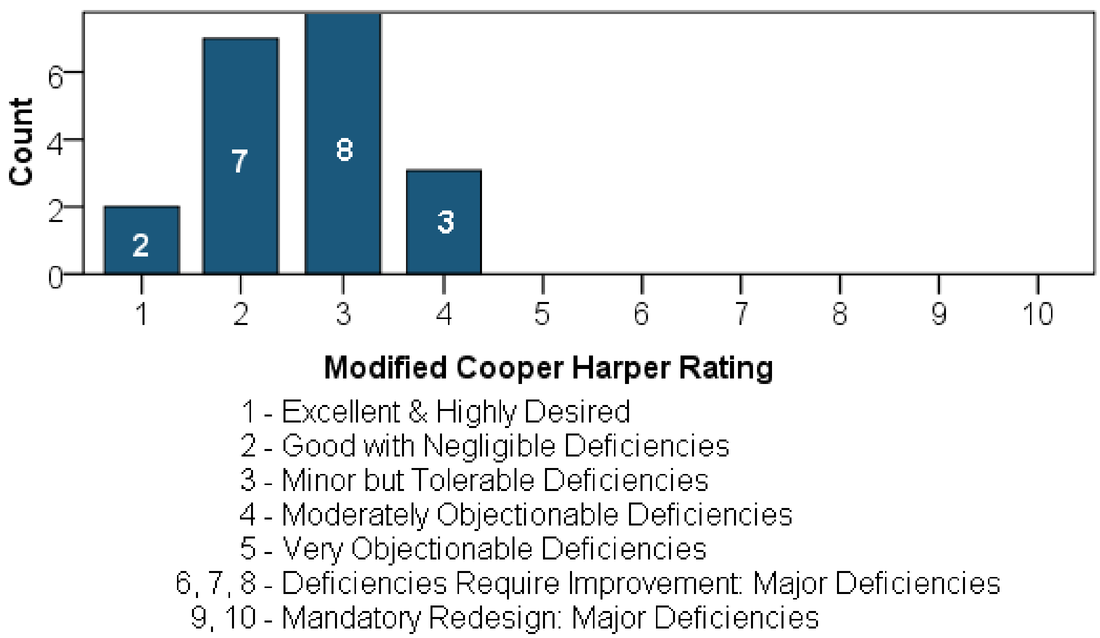


Figure 17: Modified Cooper-Harper Rating.

B. Time to Complete

As discussed in Section II, the time to complete the LPR task is a critical value, impacting elements of lunar landing, such as the fuel consumption requirement and the design of the landing trajectory. Therefore, improved understanding as to the factors driving time to complete will assist in system design and mission planning. In this experiment, there were several hypotheses regarding task completion time.

First, task modeling of LPR implied that environmental and scenario factors would affect the time to complete. The logic behind this hypotheses was that an increase in the amount of external cues (such as number of ITM, number of POI) would result in additional information processing time. In particular, the terrain expectancy factor was included in the task execution time model to account for the terrain orientation exercise performed by expert pilots. A three-way repeated measures ANalysis Of VAriance (ANOVA) was used to test the significance (α = 0.05) of POI, ITM, and expectancy. Flight experience and pilot certification were also included in the ANOVA as covariates. There was

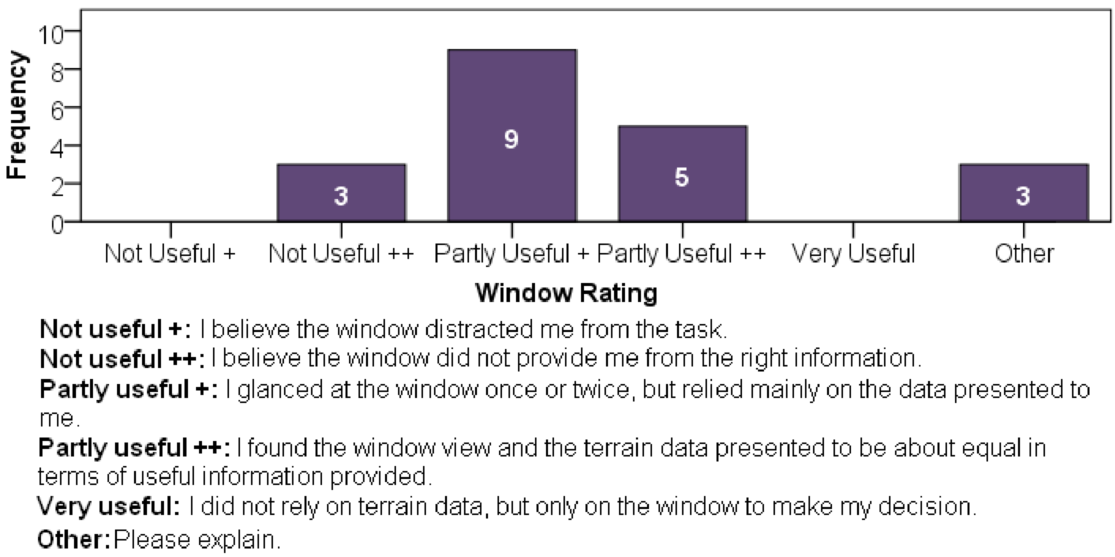


Figure 18: Window View Preference.

no significant main effect of POI on either the ITM(1,3) group, $F(1,6) = 1.372, p = 0.286$, and the ITM(2,4) group, $F(1,3) = 0.130, p = 0.742$. The main effect of terrain expectancy is not significant on either group, ITM(1,3) $F(1,6) = 0.028, p = 0.872$, and ITM(2,4), $F(1,3) = 5.315, p = 0.104$. The number of ITMs is also not significant on the ITM(1,3) group, $F(1,6) = 0.084, p = 0.782$, and the ITM(2,4) group, $F(1,3) = 0.223, p = 0.669$. The covariates, flight experience and pilot certification, were generally not significantly related to any of the main effects for either group. Table 5 lists the statistical results for each ITM group. However, the pilot certification was marginally significantly related to the number of ITMS for the ITM(1,3) group, $F(1,6) = 4.782, p = 0.071$. The flight experience was marginally significantly related to the terrain expectancy for the ITM(2,4) group, $F(1,3) = 6.460, p = 0.085$. The interaction between terrain expectancy and ITMs is reported as significant for the ITM(2,4) group, $F(1,3) = 10.349, p = 0.049$. The effects of these interactions are plotted in Figure 19.

Second, the issue of learning from previous runs was hypothesized as a potential effect on time to complete. A one-way repeated-measures ANOVA was applied to address this issue. Mauchly's test indicated that the assumption of sphericity had been violated for the main effect of run order, $\chi^2(27) = 53.120, p = 0.003$. The degrees of freedom are corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon_{GG} = 0.558$). The effect of run order was reported as not significant on time to complete, $F(3.908, 62.524) = 0.676, p = 0.608$. The covariate flight experience is not significantly related to run order, $F(3.908, 62.524) = 0.215, p = 0.926$, nor is pilot certification, $F(7, 112) = 0.860, p = 0.491$.

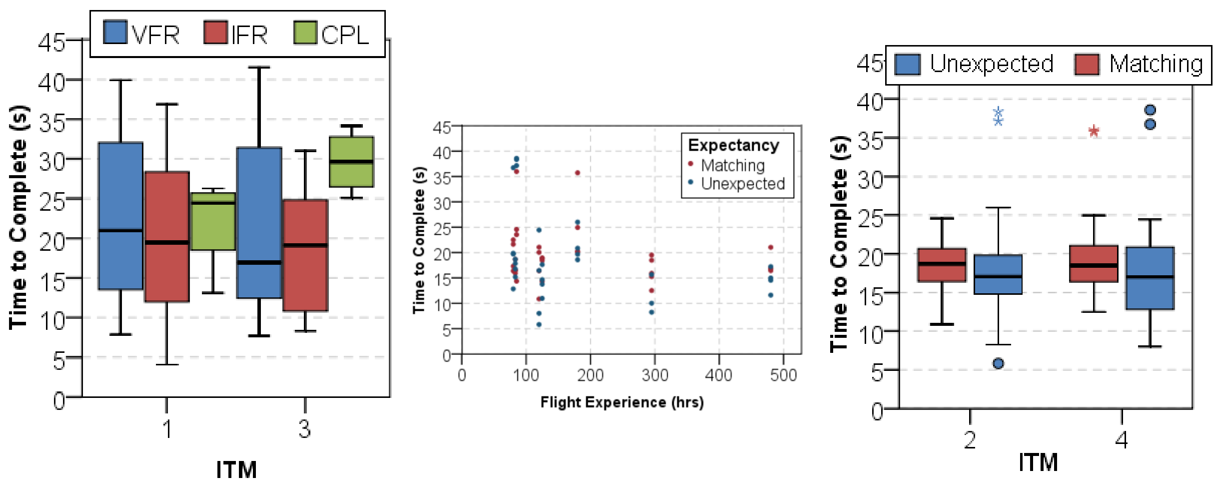
Third, the pilots were observed to perform the LPR task differently, depending on whether the run included a SAGAT questioning freeze. A two-way repeated measures ANOVA was used to test the significance of SAGAT (and the run order of the four SAGAT tests) on the time to complete. Mauchly's test indicated that the sphericity assumption has been violated for the interaction of the main effects, $\chi^2(5) = 11.857, p = 0.038$. The degrees of freedom are corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon_{GG} = 0.639$). The effect of SAGAT run order was reported as not significant, $F(3, 36) = 0.684, p = 0.568$. The interaction between the main effects are not significant, $F(1.917, 23.007) = 1.885, p = 0.150$. The covariates, flight experience and pilot certification, were generally not significantly related to any of the main effects. Table 6 lists the statistical results.

The effect of SAGAT on/off was reported as marginally significant on time to complete, $F(1, 12) = 76.719, p = 0.091$. In addition, the covariate flight experience is marginally significantly related to SAGAT on/off, $F(1, 12) = 3.228, p = 0.098$. The effect of these interactions are plotted in Figure 20.

Lastly, a two-way independent measures ANOVA was used in order to determine if flight experience and pilot certification were significant main effects on time to complete. The pilot certification was not a significant effect, $F(1, 134) = 1.069, p = 0.303$, but the flight experience was a significant effect on time to complete, $F(15, 134) = 8.692, p = 1 \times 10^{-13}$. Flight experience seems to be negatively related to time to complete, $r = -0.260, p < 0.020$, meaning that pilots of greater flight hours generally completed the task quicker than pilots of less flight experience. Figures 19band 20b illustrate this relationship.

Table 5: Results on the Impact of POI, ITM, and Expectancy on Time to Complete.

ITM(1,3)				ITM(2,4)			
Source	**df, Err. df**	F	p	**Source**	**df, Err. df**	F	p
POI $\times \epsilon$	1,6	0.053	0.826	POI $\times \epsilon$	1,3	0.308	0.617
POI $\times$ ITM	1,6	0.890	0.382	POI $\times$ ITM	1,3	0.223	0.669
ITM $\times \epsilon$	1,6	0.016	0.903	ITM $\times \epsilon$	1,3	10.349	0.049
POI $\times \epsilon \times$ ITM	1,6	0.533	0.493	POI $\times \epsilon \times$ ITM	1,3	3.463	0.160
POI $\times$ Flight Hrs	1,6	0.161	0.702	POI $\times$ Flight Hrs	1,3	2.2×10^{-4}	0.989
POI $\times$ Pilot Cert.	1,6	0.030	0.869	POI $\times$ Pilot Cert.	1,3	0.004	0.956
$\epsilon \times$ Flight Hrs	1,6	0.024	0.881	$\epsilon \times$ Flight Hrs	1,3	6.460	0.085
$\epsilon \times$ Pilot Cert.	1,6	0.432	0.536	$\epsilon \times$ Pilot Cert.	1,3	5.208	0.107
ITM $\times$ Flight Hrs	1,6	4.782	0.071	ITM $\times$ Flight Hrs	1,3	0.155	0.720
ITM $\times$ Pilot Cert.	1,6	0.612	0.464	ITM $\times$ Pilot Cert.	1,3	0.119	0.753



(a) ITM and Pilot Certification on the ITM(1,3) group

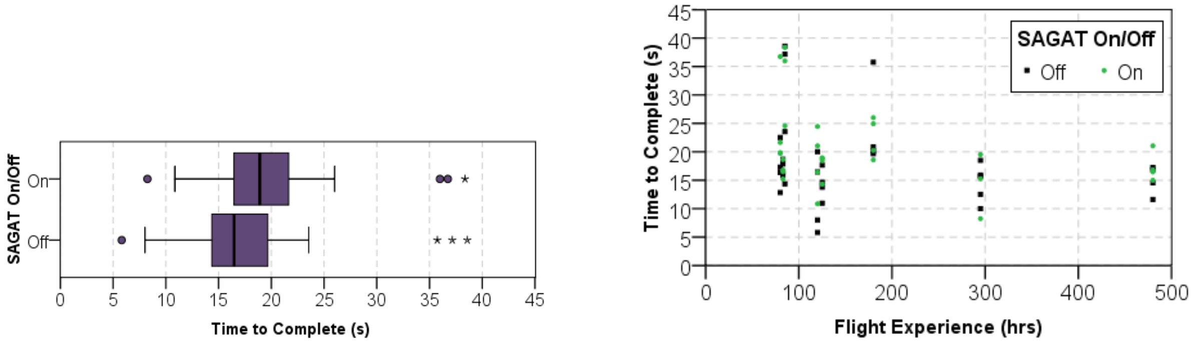
(b) Terrain Expectancy and Flight Hours on the ITM(2,4) group

(c) Terrain Expectancy and Flight Hours on the ITM(2,4) group

Figure 19: Effect of Covariates and Main Effects on Time to Complete.

Table 6: Results on the Impact of SAGAT On/Off and SAGAT Run Order on Time to Complete.

Source	**df, Err. df**	F	p	**Source**	**df**	F	p
On/Off $\times$ Flight Hrs.	1,12	3.228	0.098	On/Off $\times$ Pilot Cert.	1,12	2.171	0.166
SAGAT Run Order $\times$ Flight Hrs.	3,36	1.372	0.267	SAGAT Run Order $\times$ Pilot Cert.	3,36	0.969	0.418
On/Off $\times$ SAGAT Run Order $\times$ Flight Hrs.	1.92,23.01	0.661	0.520	On/Off $\times$ SAGAT Run Order $\times$ Pilot Cert.	1.92,23.01	1.926	0.170



(a) SAGAT on/off

(b) SAGAT on/off with Flight Hours

Figure 20: Effect of SAGAT On/Off on Time to Complete

These factors affecting time to complete, especially the lack of significant effects from POI, terrain expectancy, and ITM, are surprising. There are several plausible reasons for the lack of effect on time to complete by the scenario and environmental effects. First, there were two common approaches pilots used to handle the POI. Some pilots focused on one of the two sites, eliminating the other completely as a landing site attraction. The basis for elimination were due to hazard proximity or fuel contour location. This decision was sometimes made based on the satellite photography provided before every run, or during initial evaluation of the LIDAR sensor scan results. Other pilots would attempt to find a landing site that was equal distance from the two POI. In both approaches, pilots noted the use generating a mental map of the expected landing area and mentally highlighting/focusing on favored “sweet spots”, “zones”, or “quadrants” of where ideal landing sites were desired to occur. As such, pilots may not have spent a significant amount of time evaluating the second POI. Second, the not significant effect of terrain expectancy may be caused by multiple sources, such as inadequacy of the experimental design or varying strategies. The visuals used to simulate terrain expectancy may not have been sufficiently different (see Figure 12) to cause an appreciable delay in the LPR task. Results from the ANOVA imply that pilots of varying flight experience view terrain expectancy differently. Pilots reported not evaluating the satellite photography before receiving the LIDAR sensor scan results. However, several pilots did note that the terrain looked “different” or “unexpected”. One pilot, who was VFR rated, remarked that he disregarded the satellite photography in later runs, commenting on “needing some time to reorient and figure out what I’m looking at”. Another VFR pilot mentioned being surprised by the results of the LIDAR scan, when discovering that “[the landing area] wasn’t as bad as [he] thought it would be”. These issues associated with terrain expectancy are similar to those regarding the effect of ITMs. The ITMs, in both the experiment and the model, are quantified by discrete values. However, observations from the experiment imply that the logic enumerating the ITMs is not universal. The design of the experiment maps may not have established an appreciable difference between a low (1,2) and high (3,4) number of ITMs. For example, in 28 attempts (of all pilots) at answering the SAGAT question, “How many hazards are on the screen”, in only twelve instances was the right answer provided. Furthermore, pilots seemed to regularly underestimate the hazardous nature of the expected landing area. The experimental maps were designed to have 66% - 75% non-hazardous area. When prompted with the SAGAT question, “What percentage of the map is hazard-free? (Don’t know, $25, 33, 50, 66, 75\%$)” only 8 out of 21 attempts at this question were correct. The wrong answers were generally lower percentages or unknown. The results of the ANOVA also imply that different pilot certification (VFR vs. IFR) may approach ITMs differently, particularly when the AFM hazard identification differed from the results of the LIDAR sensor scan. In several scenarios applied to this lunar map, at least one alternative site was placed directly over the magnified region in Figure 21.

The pilots were told to handle the LIDAR sensor results as they would with their standard flight instruments, as to avoid any unnecessary bias with regards to pilot’s perception of automation fidelity. The responses of VFR and IFR pilots on the interpretation of lunar terrain geometry provided significant insight to this hazard identification process. One VFR pilot treated this region as a hazard, commenting on how he “didn’t really trust the [LIDAR] sensor” as this region was not identified as a dangerous spot. In contrast, several IFR stated they were trained to “trust [their] instruments”. One IFR pilot described the difficulty in visually finding the local horizon under sub-par conditions, as cloud layers and atmospheric effects would confound intuition. She applied these experiences to this experiment and treated the terrain as “probably just shadows”. Even though this experiment did not intentionally measure trust in

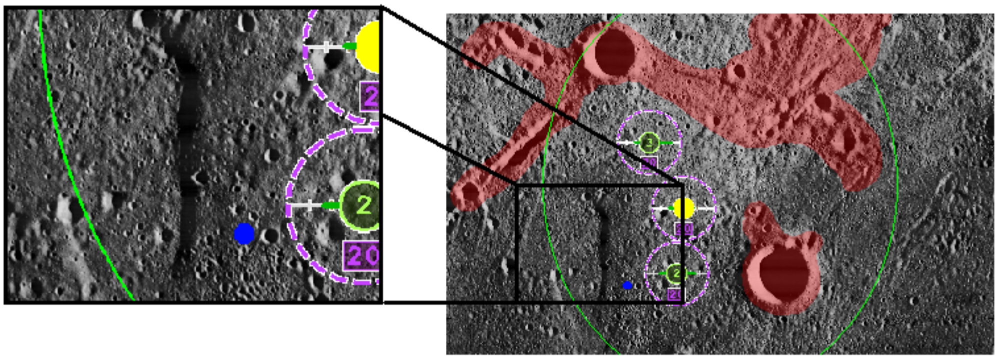


Figure 21: Ambiguity of Lunar Terrain due to Lighting and Ground Geometry Effects.

automation, this incident may be one of the contributing factors to time completion.

Although the reported results on the effect of SAGAT questioning were reported as marginally significant, experimental observations implied the effect should have been significant. Although the reported results on the effect of SAGAT questioning are not surprising, the level of impact is unexpected, given experimental observations. SAGAT is intended to unobtrusively gather information regarding situation awareness. However, this secondary task had unanticipated consequences. Common complaints noted the length of SAGAT questions (in terms of wording and comprehension of question meaning), the general interruption into a high workload scenario, and questions regarding a strategy that they did not employ. With regard to the latter, the SAGAT questions were written assuming that pilots would be able to identify any alternative site based on their AFM ranking/identification, without visual cues. However, as several pilots noted, those alternative sites were identified based on the relative position of the site to some other measure, such as within a mental projection of a four quadrant space. Thus, when asked to identify which of the four (baseline, 1, 2, 3) sites exhibited a certain terrain characteristic, the pilot was unable to recall the site number, but knew, for example, that the upper left site was in an area of favorable slope and roughness margins. The pilots also expressed frustration at interruptions to the LPR task. On several runs, pilots were observed to be momentarily confused after completing the SAGAT questions. Some pilots rapidly switched between two or three hot keys; others pressed a hot key despite engaging the landing site selection ARM button. The pilots confirmed during the debriefing that they were attempting to recall the last task prior to SAGAT questioning. One pilot believed he could have completed the LPR task faster, had the SAGAT test not interrupted his thinking. Although SAGAT tests were introduced in the practice runs, perhaps a change in mental intensity between the practice and actual experiment runs contributed to pilots' struggles in transitioning from SAGAT test to experiment.

C. Accuracy of Selected Landing Site

The success of a mission is heavily dependent on the accurate selection of a landing site. The quality of the landing site is based on Equation 5 in both continuous and ordinal form. This analysis uses both forms of the variable. As with the measure time to complete, there were several hypotheses regarding the factors affecting accuracy of selected landing site. Understanding these factors can assist in better astronaut training to counter adverse effects and can aid in collecting evidence for the role of the astronaut pilot during lunar landing.

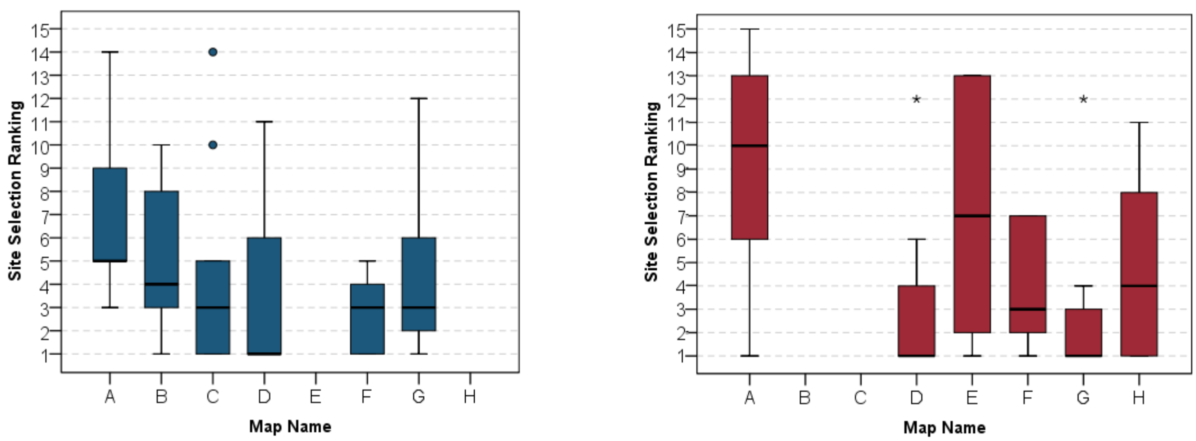
First, several pilots commented on the familiarity of some lunar maps. Although there were eight lunar maps total, each pilot saw eight different scenarios over six different lunar maps. This limited number of maps is due to the unavailability of maps with the terrain features needed in this experiment. All maps are modified images from the Lunar and Planetary Institute.[40] As such, two maps were seen twice each in the actual experiment. The unexpected scenario and the respective map was always seen before the repeat of the same map, to avoid negating the effect of terrain unexpectancy. The same map was not seen twice in a row - at least one different map separated a repeat viewing. However, the concern that familiarity with the landing terrain might cause a bias. To determine whether viewing the landing maps twice was an effect on accuracy of landing site, Friedman's ANOVA was applied to the four maps (two maps each only viewed by the two ITM groups). The results of this test are listed in Table 7. The results of Friedman's ANOVA report no significant effect on viewing a lunar map twice.

Second, a few pilots had mentioned that certain lunar maps were more difficult than others. Although the maps were designed to be equivalent and experiment trial runs did not report a map bias, pilot perceptions of their landing environment may be more tuned to smaller details. Comparing the accuracy in site selection with respect to

Table 7: Effect of Repeated Views of Maps on Accuracy of Landing Site.

	Map B	**Map C**	**Map E**	**Map H**
df	1	1	1	1
χ^2	1.286	0.333	0.200	0.200
p	0.257	0.564	0.655	0.655

each map (Figure 22) suggested there might be an inequality. Friedman's ANOVA was used to test this hypothesis. There were significant effects on the accuracy of the site selection based on the maps seen by the ITM(1,3) group, $\chi^2(5) = 12.550, p = 0.028$, and marginally significant effects from the maps viewed by the ITM(2,4) group, $\chi^2(5) = 10.944, p = 0.053$. Comparisons of the maps (Figure 22) showed that map A was more challenging than the others. To account for this nuisance variable, maps were included as a covariate in analyses relating to accuracy of selected landing site, enough degrees of freedom permitting.



(a) ITM(1,3) group.

(b) ITM(2,4) group.

Figure 22: Comparison of Map Terrain on Site Selection Ranking.

The third hypothesis concerned the impact of environmental and scenario factors on accuracy of landing site selection. The site selection ranking was expected to deteriorate as the number of POIs and ITMs increased, and especially in instances where the terrain was unexpected. A three-way repeated measure ANOVA was used to determine the significance of these main effects. The results of this test reported POI as not significant for the ITM(1,3) group, $F(1, 6) = 0.156, p = 0.706$, nor is it significant for the ITM(2,4) group, $F(1, 3) = 0.159, p = 0.717$. The effect of terrain expectancy is not significant for the ITM(1,3) group, $F(1, 6) = 0.916, p = 0.376$ and is also not significant for the ITM(2,4) group, $F(1, 3) = 1.399, p = 0.322$. The number of ITMs is marginally significant for the ITM(1,3) group, $F(1, 6) = 5.926, p = 0.051$ but is not significant for the ITM(2,4) group, $F(1, 3) = 0.265, p = 0.642$. Figure 23 illustrates the effect of ITM on Accuracy of the selected landing site. The interactions and relations with covariates are listed in Table 8.

The fourth, fifth, and sixth hypotheses, regarding the effect of run order, SAGAT questioning, and pilot SA level, were formulated based on the same observations with respect to the time to complete metric. Similarly, each additional run may significantly assist the pilots in selecting better landing sites. The interruptions to the LPR task due to SAGAT questioning may have affected the selection site quality. A one-way repeated measures ANOVA was utilized to test the fourth hypothesis. The results of Mauchly's test show that the assumption of sphericity is violated for run order. The Greenhouse-Geisser degree of freedom correction is applied (ϵ_{GG} = 0.383). The main effect, run order, is reported as not significant, $F(2.680, 21.442) = 0.712, p = 0.541$. Table 9 lists the statistical results of this analysis. No covariates were significantly related to the run order.

The fifth hypothesis was tested using a two-way repeated measures ANOVA. The main effect of SAGAT on/off was reported as not significant, $F(1, 12) = 2.354, p = 0.151$. SAGAT run order was also reported as not significantly affecting the accuracy of the landing site selection, $F(3, 36) = 0.708, p = 0.553$. Table 10 lists the statistical results

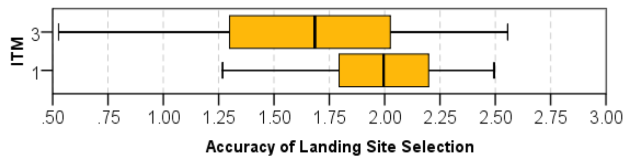


Figure 23: Effect of ITM on Accuracy of Landing Site Selection for the ITM(1,3) group.

Table 8: Results on the Impact of POI, ITM, and Expectancy (ϵ) on Accuracy of Selected Site.

ITM(1,3)				**ITM(2,4)**			
Source	**df**	F	p	**Source**	**df**	F	p
POI $\times\epsilon$	1	0.298	0.605	POI $\times\epsilon$	1	0.702	0.464
POI $\times$ ITM	1	0.317	0.594	POI $\times$ ITM	1	3.024	0.180
ITM $\times\epsilon$	1	1.948	0.212	ITM $\times\epsilon$	1	3.101	0.176
POI $\times\epsilon\times$ ITM	1	1.128	0.329	POI $\times\epsilon\times$ ITM	1	0.014	0.915
POI $\times$ Flight Hrs	1	1.827	0.225	POI $\times$ Flight Hrs	1	0.120	0.752
POI $\times$ Pilot Cert.	1	1.681	0.242	POI $\times$ Pilot Cert.	1	0.004	0.954
$\epsilon\times$ Flight Hrs	1	1.104	0.334	$\epsilon\times$ Flight Hrs	1	0.594	0.497
$\epsilon\times$ Pilot Cert.	1	3.723	0.102	$\epsilon\times$ Pilot Cert.	1	0.666	0.474
ITM $\times$ Flight Hrs	1	0.000	0.984	ITM $\times$ Flight Hrs	1	0.280	0.633
ITM $\times$ Pilot Cert.	1	0.272	0.621	ITM $\times$ Pilot Cert.	1	0.420	0.563

Table 9: Results on the Impact of Run Order and Covariates on Accuracy of Selected Site.

	Run Order $\times$		**Run Order $\times$ Maps during run**							
Source	**Flight Hrs**	**Pilot Cert.**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
df,9.564	1.027	1.609	2.451	0.826	0.657	1.017	1.320	1.812	1.336	1.583
F	0.859	1.346	2.050	0.690	0.550	0.851	1.104	1.515	1.118	1.324
p	0.467	0.285	0.142	0.552	0.635	0.470	0.364	0.241	0.365	0.256

of this analysis. No covariates were significantly related to SAGAT run order.

Table 10: Results on the Impact of SAGAT On/Off and SAGAT Run Order on Accuracy of Selected Site.

Source	**df, Err. df**	F	p	**Source**	**df**	F	p
On/Off × Flight Hrs.	1,12	1.800	0.205	On/Off × Pilot Cert.	1,12	3.313	0.094
SAGAT Run Order × Flight Hrs.	3,36	0.742	0.534	SAGAT Run Order × Pilot Cert.	3,36	0.216	0.884
On/Off × SAGAT Run Order × Flight Hrs.	3,36	0.870	0.466	On/Off × SAGAT Run Order × Pilot Cert.	3,36	0.915	0.443

The sixth hypothesis was tested by determining correlations between each level of SA and the site selection ranking. The amount of correct SA Level 1 questions was reported as non significant on the site selection ranking, $\tau = -0.062, p = 0.283$. SA Level 2, however, was a marginally significant effect, $\tau = -0.132, p = 0.093$, along with SA Level 3, $\tau = -0.217, p = 0.056$. Both of these are negative correlations, implying that increased awareness at these two levels assists in better site selection. Figure 24 illustrates these trends.

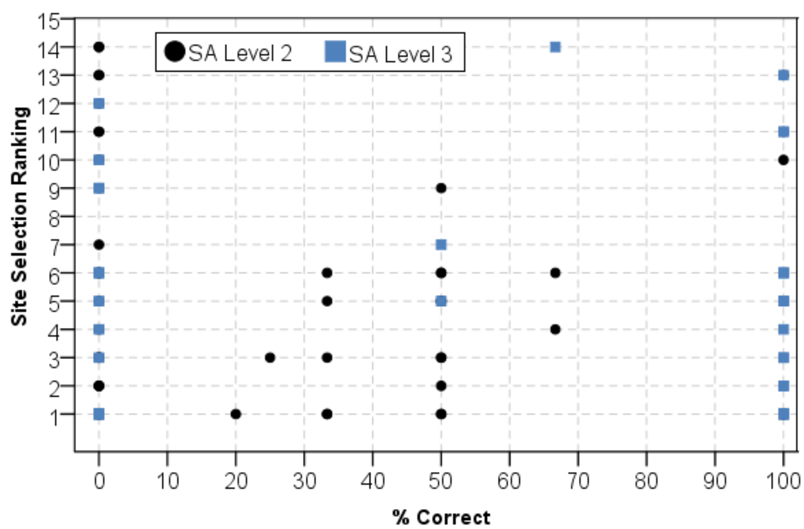


Figure 24: Trend of SA level and Site Selection Ranking

Lastly, the seventh hypothesis is with regard to the effect of flight experience and pilot certification on the accuracy of the landing site selection. The LPR task, in the most fundamental form, is similar to the in-flight emergency known as "engine out". Engine out occurs when a single-engine aircraft experiences an engine failure. At this time, the pilot needs to make an emergency landing over land.[41] As described by several pilots, the standard procedure involves stabilizing the aircraft and finding a suitable place to land. The pilot must quickly find a landing area of level and relatively smooth terrain and choose this place to land. He cannot deliberate too extensively on where to land, as time is critical and the pilot must ensure he can maneuver to the desired landing location. This procedure is a required skill for all licensed pilots. As all participants are aware of this procedure from standard pilot certification, the more experienced pilots are believed to select better sites. To test this hypothesis, a two-way independent ANOVA was utilized. The main effect of flight hours was reported as not significant, $F(15, 134) = 0.809, p = 0.666$. The effect of pilot certification was also deemed not significant, $F(1, 134) = 0.069, p = 0.561$. However, a low selection site score could also be a highly ranked site, relative to the other alternatives. To examine this trend, a non-parametric test was applied to determine the correlation between the main effects and the site selection rank. There is a significant positive correlation between the pilot certification and the site selection rank, $\tau = 0.414, p = 0.015$. Of the three pilot certifications represented in this experiment, VFR pilots generally picked better sites than IFR and CPL pilots. However ,a non significant negative correlation exists between flight experience and site selection rank, $\tau = -0.012, p = 0.423$.

The results of the statistical analysis regarding accuracy of landing site and site selection ranking are surprising. As hypothesized, environmental and scenario factors were expected to cause the accuracy of the site to decrease. Except for the ITM(1,3) group and the number of ITMs, none of these factors were considered significant. There are several possible reasons for this result. The reasons stated for time to complete (Section B) are most likely applicable here: regarding two POI as one due to strategy; inadequacies in the experimental design with respect to terrain expectancy; differences in interpretation of terrain features. The ITM(1,3) group, which is about an equal distribution of IFR and VFR pilots, tended to pick worse sites when the number of ITMs increased. This effect, which was expected, may be caused by other factors not captured in this particular set of statistical analysis.

D. Situation Awareness

The LPR task requires a substantial amount of mental workload and some degree of SA. However, the level of SA required to complete LPR, much less the necessary level for accurate LPR, is unknown. There are two aspects of SA that should be investigated. The current SA level of the pilot permits better understanding as to the utilized strategies and the associated information used for task completion. The SA level provided by the display assists in maintaining the pilot's SA and acts as an essential rapport for pilot-work domain interaction. Intuitively, a poor display design can hinder or worsen a pilot's SA, whereas a well-designed display supports and enables more efficient work. This experiment was used to answer several hypotheses regarding SA. An increased appreciation for the issues regarding SA during LPR were determined from the results of the following analyses.

First, the LPR display was designed in a manner to support SA. If designed appropriately, then the pilots should be able to answer the majority of the SAGAT questions. The answer to this question is not a succinct conclusion to the capabilities of the display, as the ability to answer SAGAT questions is dependent on the pilot's obtainable SA. However, this measure does shed some initial insight to the efficacy of the LPR display. As seen in Table 4, there were eighteen SAGAT questions. One question, #15, was eliminated from further consideration due to the ambiguity of the question. Multiple pilots did not recognize that the opening screen of the LPR task (where satellite photography of the expected landing area was provided) was considered scenario training. In total, there were seven Level 1, seven Level 2, and three Level 3 questions. Table 11 lists the number of question attempts and the percentage correct.

Table 11: Results of the SAGAT questions

	Level 1							**Level 2**							**Level 3**		
Question #	**1**	**2**	**5**	**6**	**7**	**9**	**18**	**3**	**4**	**8**	**11**	**12**	**14**	**17**	**10**	**13**	**16**
Attempts	27	22	26	29	28	19	15	21	16	23	24	22	8	17	28	17	15
% **Correct**	44	27	73	76	61	63	33	38	31	30	19	32	13	29	57	35	60
Total Correct	93/166 = **56**%							37/128 = **29**%							31/60 = **52**%		

Overall, over half of the Level 1 and 3 questions were answered correctly. The performance on the Level 2 questions was not as high, with about a quarter of the questions answered correctly. The display may provide SA Level 1 and 3 but may also be lacking in providing SA Level 2. Further investigation is necessary, to determine the true SA capabilities of the LPR display. However, these results may be influenced by other factors.

The second hypothesis concerns whether environmental and scenario factors have an impact on the different levels of SA. As a scenario becomes more complex, with multiple POI, unexpected terrain, and several ITMs, the SA levels may be negatively affected. The perception of data (Level 1) may decrease, but the comprehension of meaning (Level 2) and projection of near future (Level 3) are most likely unaffected. Three one-way repeated measures ANOVAs were used to determine the significance of the main effects. The main effect of POI was reported not significant to SA Level 1, $F(1,10) = 0.673, p = 0.431$, Level 2, $F(1,14) = 0.923, p = 0.353$, or Level 3, $F(1,8) = 0.631, p = 0.450$. The terrain expectancy did not have a significant effect on Level 1, $F(1,10) = 0.002, p = 0.968$, Level 2, $F(1,11) = 0.277, p = 0.609$, or Level 3, $F(1,8) = 0.807, p = 0.395$. The number of ITMs was not significant to Levels 1, 2, or 3, for either the ITM(1,3) or ITM(2,4) groups. Table 12 lists the results of the statistical analyses for POI, ϵ, ITM, and the covariates flight hours (FE) and pilot certification (PC). The ITM main effect was not modeled with covariates due to a lack of degrees of freedom.

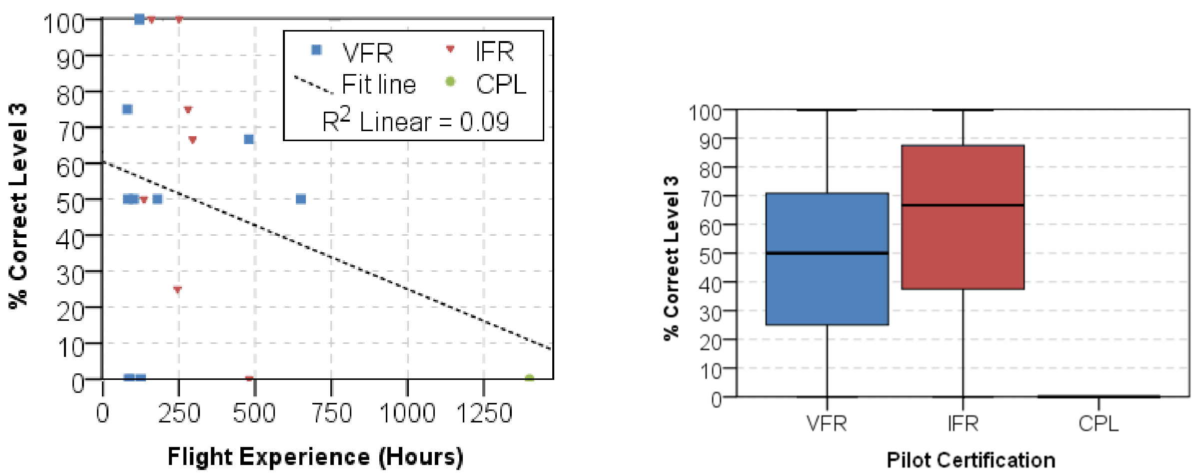
The third hypothesis relates to an observation of pilot behavior. Several pilots mentioned they were unable to correctly answer some of the SAGAT questions because they were unaware of what they would be questioned on. The SAGAT questions were designed as such that the pilots would not know all of the questions beforehand, to avoid biasing the pilot strategy. However, multiple SAGAT runs may have created a learning effect, where pilots were more capable of answering SAGAT questions in later runs. A one-way repeated measures ANOVA was used to determine whether SAGAT run order played an impact on the percent correctness of the SAGAT questions. The results of this test reported that the SAGAT run order did not have a significant effect on the number of correctly answered Level 2

Table 12: Results of POI, terrain expectancy, ITM, and covariates on SA Levels 1, 2, and 3.

	Source	ITM(1,3)	ITM(2,4)	POI × FE	POI × PC	ϵ × FE	ϵ × PC
SA Level 1	**df, Err. df**	1,10	1,6	1,10	1,10	1,10	1,10
	F	0.507	0.107	0.066	0.408	0.399	0.123
	p	0.493	0.755	0.802	0.538	0.542	0.733
SA Level 2	**df, Err. df**	1,10	1,7	1,14	1,14	1,11	1,11
	F	0.490	0.057	0.480	0.005	0.510	0.094
	p	0.500	0.819	0.500	0.946	0.490	0.765
SA Level 3	**df, Err. df**	1,4	1,3	1,8	1,8	1,8	1,8
	F	0.444	2.455	0.000	0.568	0.461	0.092
	p	0.541	0.215	0.991	0.473	0.516	0.770

questions, $F(3,23) = 0.665, p = 0.579$. There are no results for SA Level 1 and 3, due to a lack of representative data over the four SAGAT runs.

Lastly, although all pilots, regardless of flight experience and pilot certification, should experience all three levels of SA during flight, the question was posed whether more experienced pilots or pilots of varying certification were more inclined to answer the questions correctly. A two-way independent ANOVA was used to examine this question. The effect of flight experience was reported as not significant on SA Level 1, $F(15,18) = 3.868, p = 0.381$, or Level 2, $F(15,18) = 2.551, p = 0.462$. However, flight experience was reported as significant on SA Level 3, $F(15,18) = 5.38 \times 10^{28}, p = 0.000$. The effect of pilot certification was reported as not significant on SA Level 1, $F(1,18) = 0.043, p = 0.869$, Level 2, $F(1,18) = 0.000, p = 0.999$. The pilot certification, like flight experience, was reported as significant on SA Level 3, $F(1,18) = 8.24 \times 10^{28}, p = 0.000$. Figure 25 illustrates this effect for SA Level 3. Correlation coefficients were determined as a follow-up to the results of the ANOVA. Kendall's τ coefficient is given as non significant negative correlation, $\tau = -0.064, p = 0.359$, between flight experience and Level 3 correctness. There is no correlation reported for pilot certification and Level 3 correctness, $tau = 0.000, p = 0.500$. The differences between these correlations and the ANOVA results most likely are related to the applied assumptions.



(a) Flight Experience.

(b) Pilot Certification.

Figure 25: Comparison of Pilot Demographics and SA Level 3 Correctness.

The results of the statistical analysis provide insight to the factors that affect SA and the capability of the display to provide SA. However, these results are not entirely conclusive, due to a number of external factors. For example, as mentioned in Section IV.B, assumptions were made regarding the strategy and pilot perception of data. This use of AFM recommendation identification was primarily applied to Level 2 questions. Therefore, the low percentage of correct Level 2 questions (relative to the other levels) is most likely the result of this disparity. Most likely, had the questions prompted the pilots to identify sites on a map, then more correct answers would be given. This visual based approach is the method employed by Level 3 questions, a potential cause for why, counterintuitively, more projection-based questions were correct than comprehensive scenario questions.

The percentage of correct answers to certain SAGAT questions was lower than expected There were three Level 1 questions asked based on information that was experienced prior to the eight LPR runs. First, #5 asked, "How many landing sites are displayed right now on the screen (Baseline and alternatives)? [1, 2, 3, 4, Don't Know]". A similar form of this question was actually asked during the experimental briefing. Pilots were told that the baseline site and four alternatives were marked with every hot key button, except for *a priori*, where the baseline and *a priori* site # was the same. Most answers were "3 sites" or "4 sites". The second question, #7, asked, "What is the landing footprint diameter? [5, 10, 15, 20, 25 m, Don't Know]". The landing footprint diameter was also a question during the pilot briefing, and is one of the few pieces of information that stays static, at 20 m, through all of the runs. Over a quarter of the question attempts were marked "Don't Know". The last question, #6, asked, "How many points of interest are on the screen? [1, 2, Don't Know]". This information, which varies per scenario, was provided to the pilot along with satellite photography of the expected landing area. As seen in Figure 11, arrows were used to draw the pilot's attention to the number and location of the POI. The percentage of correct answers to these questions is 73, 61, and 76%, respectively. These percentages indicate that one or several issues are occurring. The pilots may have misunderstood the question, due to misreadings under timed situations, or lack of exposure to that particular question. The pilot may also have forgotten or not been trained enough on this particular simulator. Additional practice with the simulator may result in a higher percentage of correct answers, and lead to better similiarity to astronauts, who are most likely highly well-trained on a vehicle interface. Furthermore, the incorrect answers to these questions may simply be a matter of erroneous button pushes. In the case of the question regarding POI, the pilots may have mentally eliminated one POI from working memory, instead focusing on the more attainable POI. The answers to this question seem to suggest this line of reasoning, as most incorrect answers are "1 point" when the answer was actually 2 POI.

The lack of significant effect due to POI, terrain expectancy, and ITM imply that pilots' SA are robust to a variety of scenarios. This result suggests that pilots may perceive the data as a collective whole, rather than "chunks" of information containing a singular element (such as two chunks for two POI). The data also suggests that pilots may be filtering information to have better control over the overall scenario. Multiple pilots stated there was " a lot going on" and several pilots commented on "partitioning" the lunar map into areas of desired and negligible spots.

E. Landing Point Redesignation Strategies

The strategies employed by the participants fell into two distinct camps and seemed to be independent of flight experience or pilot certification. This data was collected using pilot debriefing and counting frequency of occurrences. The experiment procedure was designed with the assumption that more experienced pilots would behave more like experts and employ a particular strategy rather than less flight experienced pilots. This dichotomy of strategy was noted by Klein,[26] who commented on expert behavior with regard to decision-making. Klein initially hypothesized that experts formulated options *en masse* and carefully compared and eliminated options until a global best is determined. However, his field observations proved contrary to this hypothesis. Instead, in unexpected situations lacking a clear strategy, experts reacted to external stimuli (perception of sensory cues) and quickly formulated options until a sufficient option was determined. At that point, no other options were further considered; the decision-making process was deemed complete. The behavior in Klein's original hypothesis best describes the behavior of non-experts, defined as people lacking the necessary experience and knowledge in that particular domain or field. In relation to this experiment, the non-expert behavior was expected to manifest itself in multiple hot key presses, whereas the expert behavior was expected to correspond to a singular or a maximum of two hot key presses. When questioned about their strategy, pilots who employed the multi-button press commented on a desire to find the best landing site, attempting to remember the best site under each hot key and comparing this site to others. Pilots of the single/double button press were typically concerned with one or two criteria, such as fuel or safety. As one pilot commented, "you don't want to be running out of juice [fuel], so make a quick decision and stick with it".

However, closer examination of the distribution in strategy did not follow this trend. Furthermore, pilots slightly varied the number of button presses between runs, although each pilot generally fell into one type of strategy over the other. Figures 26 and 27 show the distribution of strategies employed through the experiment. For lack of representative data, the CPL pilot is included with the IFR pilots.

While the multiple objective change strategy is popular with both VFR and IFR pilots, IFR pilots are more likely to utilize fewer button presses than VFR pilots. There does not seem to be a trend between the types of strategy and the pilot flight experience. An argument can be made that flight experience on aircraft is not analogous to even a mock lunar landing simulation, and thus, the participants probably needed more experience with the LPR task before more closely resembling the behavior of experts. Future studies should account for this learning curve, but the experiment

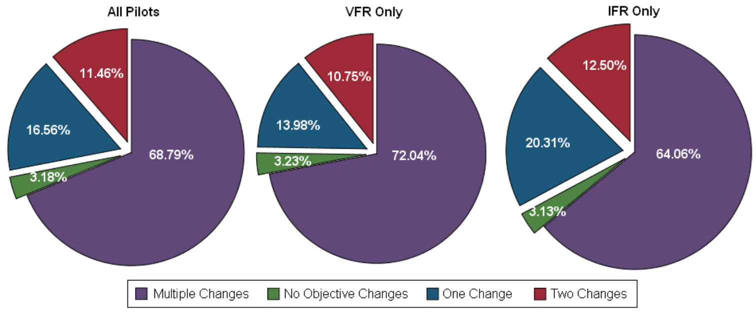


Figure 26: Distribution of Strategy Use across Pilot Certification.

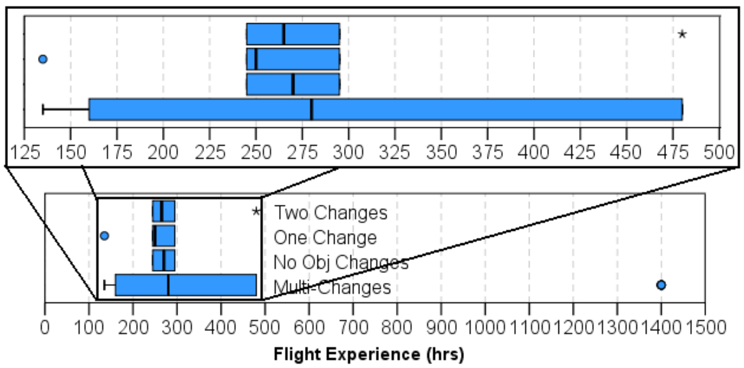


Figure 27: Distribution of Strategy Use across Flight Experience.

procedure was not originally intended to test Klein's theory of decision-making.

The pilot debriefing revealed several interested trends between the allotment of responsibility between automation and pilot, based on the pilot certification. IFR pilots seemed to be more inclined to trust the results of the LIDAR sensor scan and the AFM calculations, more so than VFR pilots who tended to doubt the validity of the LIDAR sensor results. The behavior of IFR pilots may be more consistent with expected astronaut behavior, especially since future lunar terrain is projected to occur in hazardous regions of poor visibility. Future experiments may consider recruiting IFR pilot participation only, for a stronger analogy to astronaut training. Lunar-bound crew must have some reliance on their instrument observations, especially when flying to hazardous areas; as demonstrated on Apollo 12, the viewing angle and lighting conditions can cause visual distortions on the terrain.[42] Similarly, more hazardous scenarios may prompt for an adjustable allotment of responsibility between pilot and automation, particularly in juggling objectives and information. Unexpectedly, several pilots in this study would "share" responsibilities with the simulated AFM. These pilots generally belonged to the single/double-button party, preferring to optimize on one landing metric over the others, and trusting the AFM to optimize on the other qualities. For example, the same pilot who noted the need to reserve fuel ("juice") typically engaged the balanced or *a priori* hot keys only. Based on the alternative sites calculated on those objective functions, he selected the sites that required the least amount of fuel and attempted to make his decision quickly. He typically made a decision on the landing site within 10 seconds. Since he felt fuel reserve was the most critical aspect of his vehicle, he was more confident and assured that he would be apt at optimizing fuel consumption. Other pilots exhibited similar behavior, selecting the balanced hot key and further optimizing on their number one priority. One pilot demonstrated this strategy, but her reasoning was slightly different. Her desire to save on fuel and quickly land prompted her to deliberately ignore the information presented by the other hot keys. She felt there was too much information to absorb, and thus to maximize her control over the scenario, she limited her choice of alternatives to a select few. This particular observation prompts a more in-depth discussion on automation allocation, perhaps an architecture that will adapt to the user's workload and end goals, to avoid overwhelming the user. However, this concept of automation allocation is not within the scope of this experiment.

Pilots that employed the multiple objective function change strategy also utilized various sub-strategies of searching for their desired landing site. Several pilots treated the hot key buttons as an arbitrary binning system. Each button press revealed three new choices, but as the pilots perceived them, they did not regard them as a filtering mechanism.

As one pilot stated, "[the researchers] could've labeled [the buttons] as anything, I didn't pay much attention to what they actually stood for". Because of this perception, these pilots would often view each hot key several times, refining their final decision. Generally, these pilots found the hot key arrangement to be impeditive - several pilots suggested a "see all button", a method of flagging sites for further investigation, or a means to eliminate undesired options. Although workload was not measured in this experiment, these pilots commented that recalling each site, including the hot key that revealed this site, was a challenge. All five hot keys were engaged serially on the initial review of the site alternatives; other buttons would be pressed when required for further investigation. This hot key usage strategy was not universal. Some pilots pressed the hot keys based on their preference for landing site criteria. For example, one pilot regarded the order of safety, fuel, and then proximity to POI as his preference of most important to least important. His order of hot key selection mirrored this mentality. Other pilots engaged the hot keys in the same order they were presented, but with a purpose. One pilot believed that more important objective functions should be physically closer to the pilot. He also felt that a cockpit interface designer would arrange the hot keys in a similar fashion, so pilots who were new to the vehicle (he regarded himself as such a pilot) would comprehend that this construction was the order of objective function importance. This pilot's attitude, while most likely not universal, reinforces the significance of the cockpit display interface.

The inclusion of the SAGAT questioning prompted an unintended effect on some pilots' strategies. For example, one pilot was observed to spend more time in the later runs examining the initial alternative sites, compared to his initial runs. When questioned about this peculiar change in behavior, he expressed a desire to answer more SAGAT questions correctly and as a result, spent more time examining the initial alternative sites and the location of ITMs and POIs. He stated he felt "should probably know the answers to the [SAGAT] questions". Statistical tests showed there is a marginally significant negative correlation between the percentage of correct SA questions and the site selection ranking (level 2 and 3 only). This test confirms the pilot's belief that with better SA, he was able to more confidently and accurately pick a quality landing site. Few pilots, however, exhibited this behavior. A one-way repeated measures ANOVA was used to determine the effect of run order on time to first button press. Run order was reported as not significant, $F(7, 112) = 0.706, p = 0.667$. The time to first button press for the ITM(2,4) group was not significantly affected by any environmental or scenario factors such as POI, terrain expectancy, and ITMs. However, the ITM(1,3) group was significantly affected by the interaction between POI and ITM, $F(1, 5) = 15.628, p = 0.011$. The covariates of flight experience and pilot certification related to this interaction were reported as significant, $F(1, 5) = 11.574, p = 0.019$ and $F(1, 5) = 11.189, p = 0.020$, respectively. Figure 28 and 29 illustrates these trends. For purposes of this analysis, CPL is included with the IFR group.

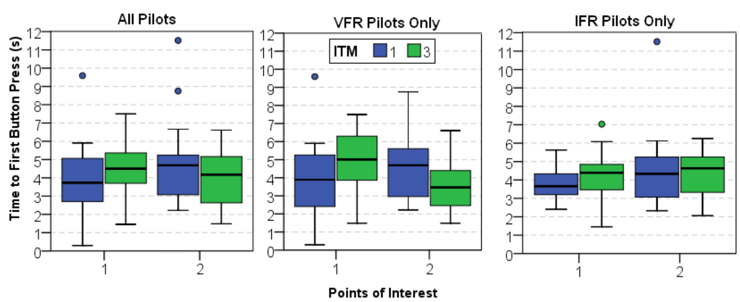


Figure 28: Effect of POI, ITM interaction with Pilot Certification on Time to First Button Press.

The pilots were also asked informally during the debriefing their mental perception of the task time remaining. Some pilots performed the LPR quickly (15 seconds or less) whereas other pilots took more time to complete the task (35 seconds or more). One pilot, who consistently used most of the 45 seconds allotted commented that he typically had a particular landing site in mind within the first twenty seconds, but would use the rest of the time to refine this site selection. When the ten second countdown warning alarm went off, he was not particularly concerned about completing the task in time, as he had his default site selection to rely on. Other pilots, however, were observed to react to the warning alarm, physically moving their hand and selecting a site within a few seconds after the first alarm chime. No pilot reported or was observed to consistently refer to the mission clock or LPR timer, and the majority of the pilots felt they had a strong sense of how long they took to complete LPR. However, the answers to the time related SAGAT questions reveal that pilots may not have been as acutely aware of the elapsed time during the actual process. SAGAT questions #2 and #18 asked, "How much time is remaining for Redesignation/Touchdown? [60-51, 50-41, 40-31, 30-21, 20-11 sec]", respectively. These two questions were answered 33% and 27% correctly out of all

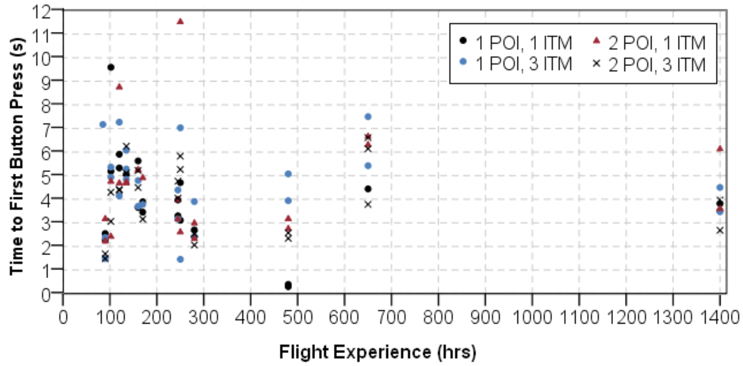


Figure 29: Effect of POI, ITM interaction with Flight Experience on Time to First Button Press.

attempts. Pilots generally overestimated the elapsed time, believing there was less time remaining for redesignation and touchdown. However, about a quarter of the answers were "Don't Know", revealing that the countdown timers were not explicitly remembered by the pilots.

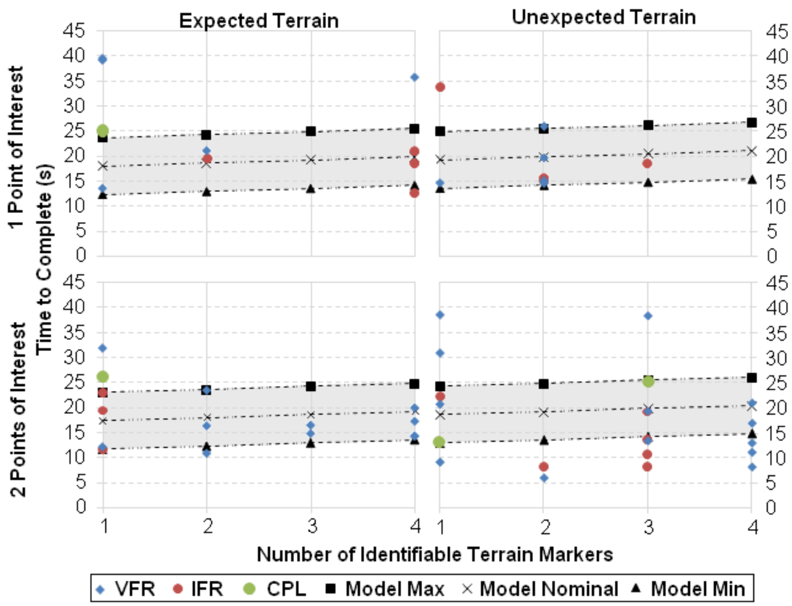
Lastly, Klein's theory states that different strategies may yield different results.[26] Since experts are not particularly concerned with finding the global best, the overall quality of their selection is generally sufficient. However, non-experts may not fare any better, as they opt to find the global best with little knowledge about the subject domain. To test this hypothesis and see if subscribers of one strategy performed better than others, a correlation was sought between the type of strategy used and the site selection ranking. The results of this positive correlation are not significant. Spearman's correlation coefficient is given as $\rho = 0.009, p = 0.472$. Therefore, for this experiment solely, multiple button presses may lead to worser site selection rankings, but this trend is not significant.

F. Model Comparisons

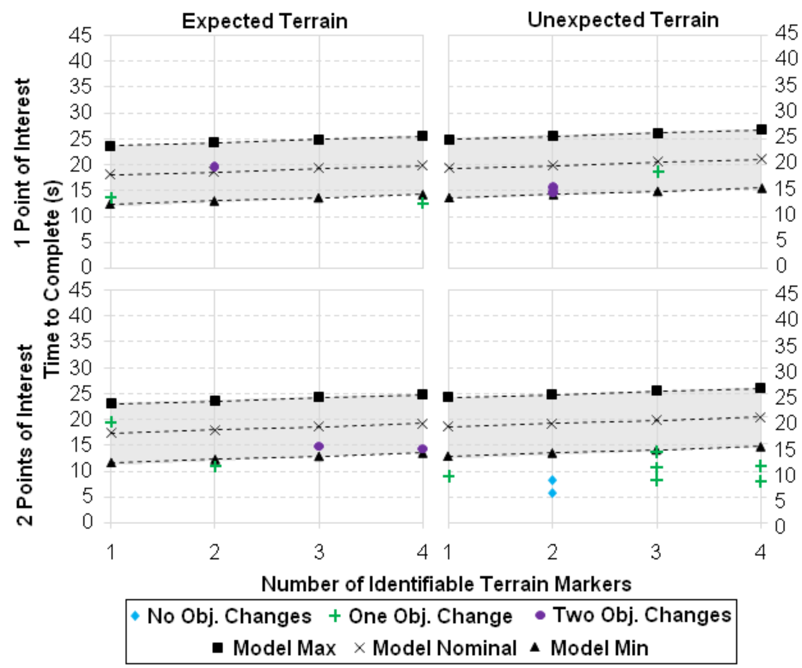
A major objective of this experiment was to inform the LPR task model based on performance analogous to lunar-bound astronauts. The LPR task predicts the time to complete, assuming the site selected is the top ranked site. This time estimation changes depending on the environmental and scenario factors of POI, terrain expectancy, and number of ITM. The model was compared against the actual results of all the number one ranked sites, across all pilot certifications. Figure 30a illustrates these results.

Overall, the model was able to fit 28 of the 53 applicable experimental data points, or 53%, within the predicted time range. The rest of the data fell almost equally below and above this time range. This fit was approximately the same for the three pilot certification groups. Of the 34 applicable data points from VFR pilots, 56% were fitted (equal distribution falling above and below); of 17 points from IFR pilots 53% of the data were fitted (the model overpredicted 35% of all data); and of the four points from the CPL pilot, 50% of the data were fitted (the model underpredicted the other half of the data). The model was also compared against experimental data that matched the same strategy assumed for the model (zero to two objective function changes) and had achieved the top site selection ranking. Figure 30b is a plot of these results.

The model did not fit this set of data as well, with only 44% of the 18 applicable data points fitting. The model tended to overpredict the time required to complete LPR, with the rest of the experimental data falling below the predicted time range. The trends presented in Figures 30b and 30b imply that several factors may be in play. First, pilots may not be behaving similar enough to astronauts, and thus this experimental data is not representative of the conditions stipulated by the model. Second, the model may not be accurately portraying the true strategies utilized for LPR. This model assumes that all subtasks are completed linearly. However, comments from the pilots imply that several tasks are completed in parallel. As seen in Section E, pilots often neglected elements of information, or examined information on a whole basis, rather than individual components. The model assumes each aspect of the scenario is individually accounted for, thus accounting for steps that are not truly executed. This model must be improved to account for these actions or changes in strategy. This deterministic model requires the use of probabilistic modeling, to provide more use in approximating the likelihood of a particular outcome or strategy use. Understanding the time to complete LPR allows for better trajectory design. While a major lander system re-design is highly unlikely, as the astronauts will most likely be trained to complete the LPR task within the constraints of the system, appreciating the impact of human control on the lunar descent and landing flight envelope can allow for further optimization



(a) Pilot Certification.



(b) LPR Strategies.

Figure 30: Comparison of LPR Task Execution Time Model.

trajectory and support system design.

V. Discussion

The analysis of the experimental results have provided significant insight to human performance during LPR. The three major experiment objectives have been obtained. First, the task model was demonstrated to predict the task completion time for about half of the applicable experimental results. The task model, however, may require a redesign to take into account different modeling strategies, such as multiple button presses. The task model can also be improved by accounting for actions done in serial and the perception of terrain data by expert pilots. Mission designers may find the task model more useful should the model provide probabilities, rather than simply discrete information. The time to complete and quality of the landing site were both fairly robust to environment and scenario factors. Learning did not significantly occur during the actual experiment and secondary tasks such as answering SAGAT questions did not significantly impact pilot performance. Second, the model was determined to overpredict the time to complete LPR for participants who utilized the same strategy of zero-to-two objective function changes. The model was not more accurate for one pilot certification or the other, as the model equally over- and underpredicted task completion time. Third, the experiment enabled the opportunity to determine the efficacy of this LPR display design from both MCH ratings and open-ended pilot feedback. The pilots generally liked the LPR display design. The LPR display provided them more information than the window view and was also able to provide them some amount of situation awareness. Incorporating pilot feedback into the second round of display design will most likely further improve overall pilot performance.

A. Improvements to the LPR Display

There were several aspects of the LPR display that the pilots liked, and other aspects that warrant improvement. Most pilots commented on the usefulness of simple symbols and relative information (in comparison to quantitative numbers). The relative information allowed for quicker processing. Absolute values were rarely provided (except in the instance of landing footprint diameter) - the pilots were unaware of the specific kilograms of fuel, or the euclidean distance between the landing site and points of interest. The overlay of relative information over the expected landing site prompted for an efficient analysis of prime sites. Almost all pilots reported looking for “long arrows [level and smooth terrain]”, “sites closer to the blue dots [near points of interest]”, “away from red areas [distance from hazards]”, and “closer to the center [of the fuel contour, implying relatively less consumption of fuel]”. Some pilots liked the inclusion of the baseline point and its accessibility from every aspect of the LPR task. One pilot found the baseline point as a useful comparison “as to what [the mission planners] thought was a good site”.

The feedback on the AFM ranking/site identification number was neutral - most pilots did not use this piece of information. The pilots reported liking the idea of having a numerical output of the landing footprint diameter, but stated they did not actively use that knowledge in the landing site selection process. Most pilots suggested a different layout of the hot keys. Pilots employing the multiple objective function strategy generally desired a “see all” button as discussed in Section E. Other suggestions regarding hot keys included a means to store or collect sites for further evaluation, or a method to clear sites no longer under consideration. One pilot suggested increasing the number of alternative sites from three to five.

The most common suggestion on display change was with respect to the slope and roughness margins. Adapted from Needham (who utilized a four axis system), this two axis system was “difficult to see”, especially when superimposed on low contrasting map images. Pilots suggested a greater use of color, with red alerting below tolerance margins, yellow signaling within tolerance margins, and green designating ideal sites. One non-pilot, who assisted in the qualification of the LPR system, advocated the use of more intuitive symbols. Instead of lines, the slope margin could be represented by some angular deflection of that axis about the local horizon. A steeper slope would cause this arrow to be at a higher angle. Roughness could be represented by a sine/cosine relation - higher amplitudes and larger frequencies would imply more rough terrain, whereas smooth terrain would tend toward a straight line. The most likely explanation for differences in response to Needham’s proposed iconography is due to participant age. Needham’s participants were generally graduate students (approximately $<$ 28 years of age), whereas pilots in this study ranged from college undergraduates to retirees. As one might expect, the majority of seeing complaints came from older pilots, where eye degradation can be a significant factor.

B. Review of Simulation and Experiment Setup

There were several areas of possible improvement to the simulation and experiment setup. First, the pseudo-AFM algorithm should be improved to better account for the slope and roughness margins of the alternative landing sites. The pseudo-algorithm could also better account for the point of diminishing returns. Several pilots noticed the algorithm's lack of sensitivity to both these issues and commented on not being able to trust the outputs for each hot key. One pilot was particularly adamant about his lack of trust in the AFM algorithm, instead choosing to search under each hot key and find the site of his liking. A greater source of equatable lunar maps and scenarios would have improved the variety of runs in this experiment. Although seeing a map twice was shown to not have a significant effect on site accuracy, a greater assortment of maps most likely would increase the possibility of independent sampling. The SAGAT questioning system could have been set up in a more uniform manner, with preplanned comparisons and contrasts of SAGAT on and SAGAT off runs. The questions should also better mimic the perception and storage of knowledge, to avoid confounding lack of SA with question ambiguity. A more equal distribution of questions at each level would help ascertain a sufficient sampling of data, as well as a longer time to answer questions. Lastly, several elements of the LPR display should change with time to more realistically mimic the LPR scenario. The fuel contour was held static in this experiment and represented the reachable area should the entire 45 seconds be used. Ideally, this contour would shrink with time, and sites would disappear as viable options. The landing footprint diameter readout should also change with time, to incorporate the continuous update of uncertainties in navigation and vehicle state. The quality of the lunar map should also change in time, to better reflect a continuous LIDAR scan.

C. Extrapolation of Experimental Results

This experiment has increased overall knowledge of human performance during LPR. This experiment was not designed to investigate or compare human control versus automated/autonomous control. However, given the results of this study, one can extrapolate and draw some initial observations on the overall topic. The trajectory of a fully autonomous landing is not modeled in this experiment. Most likely, a fully autonomous trajectory would not follow a similar trajectory design of a manned lander, as LPR would not be necessary. However, analogous automated performance could be defined as a selection of the baseline point in zero seconds (in other words, no alternative sites were considered). Performing the LPR task in zero seconds incurs no additional fuel consumption. Already, the astronauts, who require at least several seconds to absorb information and make a decision, are penalized for using fuel. However, an astronaut can potentially make better decisions as to whether a better landing site exists.

Taking this definition of analogous automated performance into account, the performance of each of the 157 total cases was compared and determined as to whether the human scored equal to or higher than the automated score. A 10% margin is applied - falling within this margin is equatable to the automated score. This comparison of true human performance takes into account the time to complete LPR and the quality of landing site selection, based on an equal weighting distribution between fuel, safety, and proximity to POI. Figure 31 summarizes the result of this comparison.

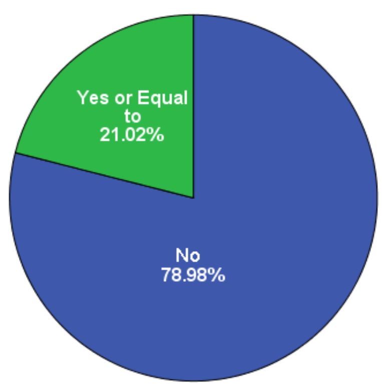


Figure 31: Comparison of Actual Pilot Selection vs Analogous Automation Selection.

The human pilots were able to select equal to or better sites than what could be achieved with an autonomous landing in 20% of all the cases. While this percentage seems low, the percentage of cases scoring higher than the analogous autonomy would most likely increase if the pilot decisions were made faster. To explore this concept, the time to complete for each case is varied and compared against automation. Figure 32 illustrates several comparison of varying, actual, and matching automated and human performance.

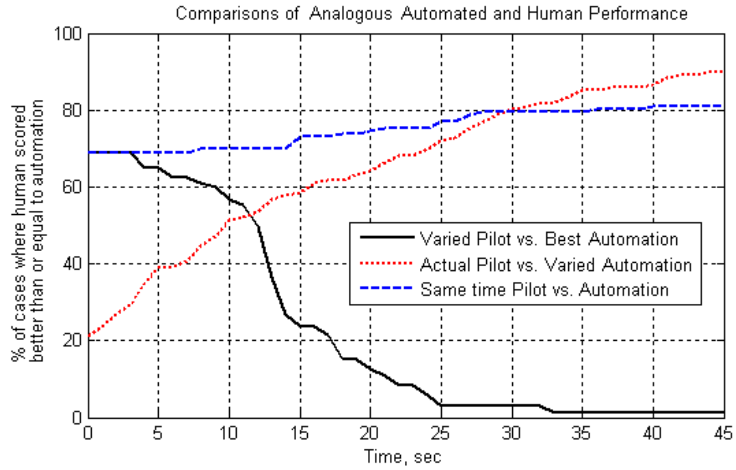


Figure 32: Comparison of Varying, Actual, and Matching Analogous Automated and Pilot Performance.

For this comparison, the varying pilot performance assumes that in each case, the pilots make the same site selection, however a hypothetical task completion time is used. This situation assumes that all cases are performed in the same amount of time. This line illustrates that if all pilots could complete the LPR task in 15 seconds or less, then 20% of all cases will score better than the best autonomous performance. The best automated performance is the standard baseline selection in zero seconds. However, beyond 25 seconds the human performance is heavily penalized by the fuel consumption, with less than 10% of the cases scoring better than the autonomy. To further emphasize the significance of completing the LPR task quickly, the actual pilot performance (site selection choice, time to complete) was compared against an autonomous lander that selects the baseline site in a variable time to complete. This graph underscores the significance of fuel consumption - after 15 seconds human performance scores higher than variable autonomous performance in 60% of all cases. Lastly, the autonomy and pilots were compared assuming both parties completed the LPR task in the same amount of time. Figure 32 clearly illustrates the advantage of performing the LPR task quickly. If the pilot can make the same quality of a decision within 12 seconds, then about half of his runs will be better than if the lander was flown autonomously. Ideally, this graphic should demonstrate asymptotic behavior around zero seconds, but in the actual experiment there were cases where the pilot picked a worse site than the baseline.

Figure 33 illustrates this comparison from a different perspective. The safety and proximity to POI scores for the baseline site are only considered, plotted against the actual time to complete for each case. As this figure illustrates, when fuel is not a significant factor, the pilots generally pick better sites than the automation. Even when expending the full allotment of fuel (45 seconds), half of the pilot site selections are equal to or greater than the automated choice. However, the percentage of better performance begins to degrade as the time to complete increases.

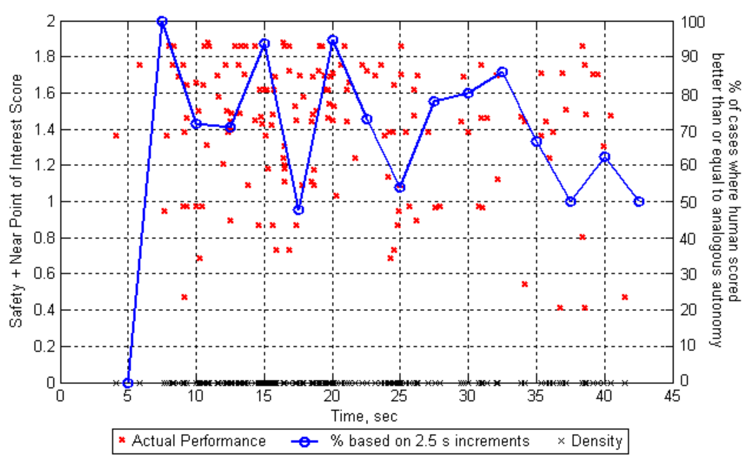


Figure 33: Comparison of Analogous Automated and Pilot Performance, Safety and Proximity to POI Scores Only.

To further examine this design space, the performance algorithm, namely, the distribution between the weighting of the three main criteria, fuel, proximity to POI, and safety, was varied from 0 (no importance) to 1 (full importance). One metric was varied at a time and the two remaining metrics were equally normally weighted based on the complement. This weighting distribution was then used in Equation 5 to calculate the final site selection score. Figure 34 illustrates contours of these weighting distributions.

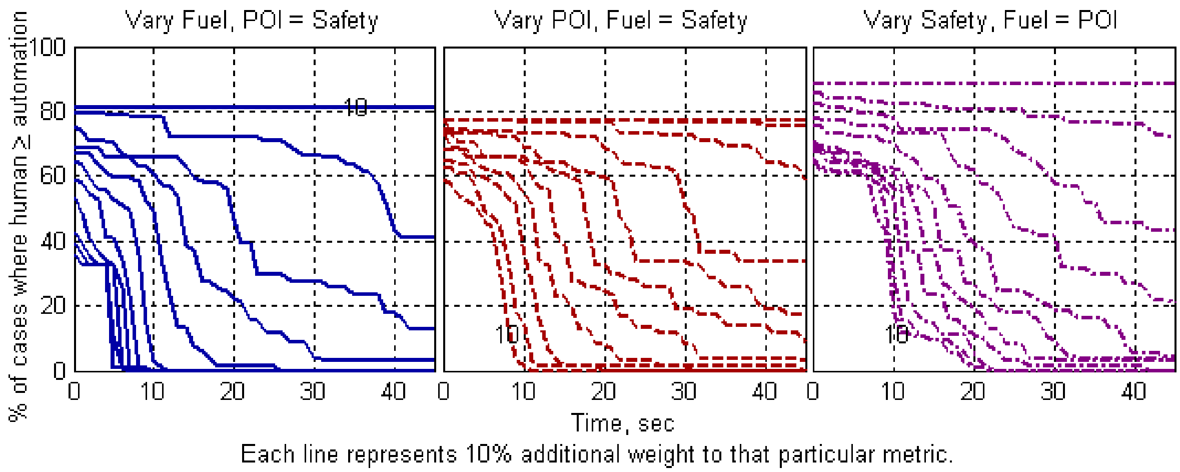


Figure 34: Comparison of Analogous Automated and Pilot Performance, Safety and Proximity to POI Scores Only.

These contours are based on actual site selections, but with hypothetical task completion times. In general, the human outperforms the automated vehicle, however, the percentage of better performance varies depending on the weighting distribution. In many ways, these contours are analogous to cumulative distribution functions (CDF). If the 157 cases performed in this experiment are representative of typical site selection choices during LPR, then these contours can assist in determining the conditions for human control. For example, if the lunar vehicle was constrained to hold enough fuel for 20 seconds of LPR, then the mission designer can compare human and automated performance for different combinations of weighting distributions. These contours show that if the primary driving factors are safety, proximity to POI, or both, at weighting distributions of at least ≥ 0.7, 0.6, or 0.45 each, respectively, then the human will choose a better landing site than a automated vehicle 60% of all cases. While these contours are useful in providing initial estimations to the LPR task, one should apply this information with the acknowledgement that the modeled human performance may not be accurately representative of true astronaut behavior and that a automated lander would most likely not operate under the same conditions as a manned lander. Thus, true CDF plots will be more useful in examining this design space.

VI. Extensions to Mars

The expected results of this study could be extended to applications of human spaceflight on Mars where the exploration of human support systems and the impact of human performance on space systems design is of great importance. The dual investigation of landing large payloads while understanding the necessary requirements to support a range of human control will assist in the development of robust spacecraft systems.

The lunar LPR task is unlikely to translate to a single opportunity during Mars EDL, unless improved technology increases the amount of rotational and translational control of the Mars vehicle. More likely, Mars-LPR will occur relatively early in the mission timeline, but require the same brevity in task completion, without sacrifice to accuracy. As such, knowing the duration of LPR (even if applied to a different terrain) will assist in understanding the required vehicle state and adjustments to the EDL event sequence. The examination into the potential human-automated trade space discussed in Section V.C illustrates these tradeoffs associated with time. Adjustments must be made to account for Mars-specific mission objectives. Furthermore, the recommendations for display design are also relevant and can be applied to landing humans on Mars. While the information presented may change to account for differences in environment (such as winds, ground temperature, etc), the lunar display provides an initial reference for humans to Mars. Lastly, the investigation of pilot population and accuracy of selected landing site provides quantitative analysis for the selection and training of future Mars astronauts.

VII. Future Work

Although the experiment results shed insight to human performance during the LPR task, additional studies are required to further the fidelity of this model. The task model is currently a deterministic model. A probabilistic model would allow examination of the full depth and breadth of human control while establishing a likelihood of outcome. This probabilistic model should also take into account steps and procedures that could potentially be completed in parallel, or the possibility of skipping subtasks during the LPR. The probabilities for these astronaut tasks will be based on the astronaut's performance shaping factors; the vehicle system state and navigational error; and the mission status. A model of this fidelity, to the author's knowledge, does not currently exist for LPR.

This model should also be used in attempt to quantify the impact of human control. Although the LPR task is one out of several anticipated crew obligations, the risk and uncertainty involving the LPR task is significantly greater than standard procedures such as system monitoring. Thus, the LPR task should be developed into a computational human performance model (CHPM). This CHPM will be integrated into a 6 degree-of-freedom (DOF) landing simulation. The final integrated product is, essentially, a human-in-the-loop (HITL) trajectory simulation. This trajectory simulation should model both translational and rotational dynamics of a pinpoint surface landing. Furthermore, this simulation must include atmospheric and gravity models, terminal guidance algorithms, terrain and inertial navigational sensors, and control actuators based on the reference vehicle design. The development of a separate landing simulation is not a trivial task, and existing 6 DOF lunar simulations should be considered as baseline reference trajectories.

Once the LPR task model is integrated with a trajectory simulation, a Monte Carlo analysis and a Design of Experiments could be used to examine a range of environmental conditions and the extent of human control. Results and observed trends from these simulations will be analyzed and summarized. Preliminary results from the process of setting probabilities and development of the CHPM will also provide insight into the human control impact on trajectory design. Elements of these determined environment states establish initial approximations on desired flight conditions. A post-run analysis of simulation results can provide full mission profiles of trajectory parameters and useful metrics. These metrics include, but are not limited to fuel consumption, time to landing, the flight envelope, and the trajectory profile. These mission profiles will be compared to profiles of autonomous missions, thus spanning the full range of human control.

VIII. Conclusion

Increasing mankind's presence in space, especially on the Moon and subsequently Mars, requires a significant amount of technological development and further understanding in the areas of human control and spacecraft design. In particular, critical mission phases such as entry, descent, and landing poses unique questions as to the role of the onboard crew and the systems necessary to support variable levels of involvement. The task of lunar landing point redesignation is an ideal case study for examining these issues. In this investigation, an experiment has been developed to inform an LPR task model, to compare true performance to the predicted performance, and to determine the efficacy of current LPR display design. The experimental results have highlighted several key conclusions regarding the LPR task.

First, the strategy assumed in the LPR task model does not match true performance. The method of LPR task execution is not necessarily linear, with tasks performed in parallel or neglected entirely. IFR pilots are also more likely to change objectives faster than VFR pilots. The elapse in time during LPR is generally not known or overestimated. Second, the time to complete the LPR task is generally robust to environmental and scenario factors such as number of points of interest, number of identifiable terrain markers, and terrain expectancy. Aspects of the expected landing area are typically regarded as an entire entity rather than individual chunks of information. Pilots of greater flight experience typically complete the LPR task faster than pilots of less flight experience. Pilots tend to underestimate the hazardous nature of the expected landing area. The accuracy of the selected landing site is also fairly robust to environmental and scenario factors. Increased levels of situation awareness generally improve the accuracy of the site selection, particularly levels of greater data comprehension and projection of near future events. Third, environmental and scenario factors do not significantly effect the amount and level of situation awareness. The perception of data may be poor due to strategy and inexperience with the display interface. Fourth, the display design provides all levels of situation awareness but the exact amount requires further investigation. The layout of the LPR display was generally well received, but the presentation of terrain characteristics should be modified to incorporate color and more intuitive markings. Pilots also suggest that additional methods of landing site comparisons be provided with this display. Fifth, the examination of the overall tradespace between the three main criteria of fuel consumption, proximity to points of interest, and safety when comparing human and analogous automation behavior illustrates that automated landing tended to perform better than actual pilot performance. However, human performance begins to excel automation performance when the time to complete decreases. An initial exploration into variable weighting of the three criteria shows that humans outperform automation in missions where safety and nearness to points of interest are the main objectives, but perform poorly when fuel is the most critical measure of performance. Lastly, the results of this exploration and can be applied to manned Mars entry, descent, and landing, and the Mars equivalent of landing point redesignation, but additional analysis is needed to draw accurate conclusions. Improvements to the fidelity of the model can be made by transgressing from a deterministic to probabilistic model and incorporating such a model into a six degree-of-freedom trajectory simulator.

This research has provided additional information regarding human performance during the landing point redesignation task. While this investigation does not fully and extensively quantify the impact of human performance on support system design, these results increase evidence to make informed decisions on larger mission design questions such as the role of humans during entry, descent, and landing. While some may view returning to the Moon as simply stepping in two-decade old footprints, this return and the subsequent set of lunar steps are providing the foundation for leaps and bounds into the next era of manned spaceflight.

Acknowledgments

The authors acknowledge the entire ALHAT project team consisting of outstanding and distinguished members from NASA's Langley Research Center, Jet Propulsion Laboratory, Johnson Space Center, C.S. Draper Laboratory, and the Applied Physics Laboratory. The program is led by Chirold Epp who deserves special thanks for his leadership and vision. The authors also thank the following individuals for their additional contributions to the project: Laura Major for her mentorship and guidance in the initial stages of this project; Robert Hirsh for his mentorship and wealth of supplementary resources; Gavin Mendeck for his advice on industry standards; Kent Lewis for his encouragement and sharing his experience with commercial aviation; James Tolbert and Vidur Vohra for their assistance in the development and execution of the LPR hardware/software simulation and experiment. Lastly, the authors thank the members of the Space Systems Design Laboratory and Cognitive Engineering Center for their continued support and assistance during this study.

References

[1] Agency, X. N., "China First Manned Space Mission," Online., 2003, Accessed May 18, 2009.

[2] Jayaraman, K., "India's Space Agency Proposes Manned Spaceflight Program," Online, November 2006, Accessed 18 May 2009.

[3] "The Vision for Space Exploration," Tech. Rep. NP-2004-01-334-HQ, National Aeronautics and Space Administration, Washington, DC, February 2004.

[4] Drake, B. G., "Reference Mission Version 3.0, Addendum to the Human Exploration of Mars: The Reference Mission of the NASA Mars Exploration Study Team," Tech. rep., National Aeronautics and Space Administration, 1998, Available Online: http://ares.jsc.nasa.gov/HumanExplore/Exploration/EXLibrary/docs/MarsRef/addendum/index.htm.

[5] Braun, R. D. and Manning, R. M., "Mars Exploration Entry, Descent, and Landing Challenges," *AIAA Journal of Spacecraft and Rockets*, Vol. 44, No. 2, March-April 2007, pp. 310–323.

[6] Sostaric, R. and Rea, J., "Powered Descent Trajectory Guidance and Some Considerations for Human Lunar Landing," *AIAA Guidance, Navigation, and Control Conference and Exhibit*, No. AIAA-2005-6287, 2005.

[7] Sheridan, T. B., *Telerobotics, Automation, and Human Supervisory Control*, MIT Press, Cambridge, MA, 1992.

[8] Bainbridge, L., "Ironies of Automation," *Automatica*, Vol. 19, 1983, pp. 775–779.

[9] Smith, P. J., McCoy, C. E., and Layton, C., "Brittleness of the Design of Cooperative Problem-Solving Systems," *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 27, 1997, pp. 360–371.

[10] Bennett, F. V., "Apollo Lunar Descent and Ascent Trajectories," Tech. rep., National Aeronautics and Space Administration, Houston, Texas, March 1970, Also presented at AIAA 8th Aerospace Sciences Meeting, January 1970.

[11] McCandless, J., Hilty, B., and McCann, R. S., "New Displays for the Space Shuttle Cockpit," *Ergonomics in Design: The Quarterly of Human Factors Applications*, Vol. 13, No. 4, 2005, pp. 15–20.

[12] Holland, A. and VanderArk, S., "Task Analysis of Shuttle Entry and Landing Activities," Tech. rep., NASA Johnson Space Center, March 1993, NASA-TM-104761.

[13] Braun, R., "Planetary Entry, Descent, and Landing Shortcourse Notes," .

[14] Steinfeldt, B., Theisinger, J., Korzun, A., Clark, I., Hutchings, A., Chua, Z., Grant, M., and Braun, R., "Human Exploration Aerocapture and EDL System Architecture Assessment," *AIAA Space 2009 Conference*, 2009.

[15] Christian, J. A., Wells, G., Lafleur, J., Verges, A., Lewis, C., and Braun, R. D., "Extension of Traditional Entry, Descent, and Landing Technologies for Human Mars Exploration," *Journal of Spacecraft and Rockets*, Vol. 45, No. 1, January 2008, pp. 130–141.

[16] Brady, T., Schwartz, J., and Straube, T., "Operational Concept Description for the Autonomous Landing and Hazard Avoidance Technology (ALHAT)," Tech. Rep. ALHAT 2.0-002, National Aeronautics and Space Administration, 2006.

[17]Davis, J., Striepe, S., Maddock, R., Hines, G., II, S. P., Cohanim, B., Fill, T., Johnson, M., Bishop, R., DeMars, K., Sostaric, R., and Johnson, A., "Advances in POST2 End-to-End Descent and Landing Simulation for the ALHAT Project," *AIAA/AAS Astrodynamics Specialist Conference & Exhibit*, No. AIAA 2008-6938, 2008.

[18]Chua, Z. K., Major, L. M., and Feigh, K. M., "Modeling Cockpit Interface Usage During Lunar Landing Redesignation," *International Symposium of Aviation Psychology*, 2009.

[19]Kessler, L. J. and Forest, L. M., "Human-Interactive Autonomous Flight Manager for Precision Lunar Landing," Presentation, Fall 2006, at AAAI Fall Symposium on Spacecraft Autonomy.

[20]Forest, L., Kessler, L., and Homer, M., "Design of a Human-Interactive Autonomous Flight Manager (AFM) for Crewed Lunar Landing," *AIAA Infotech@Aerospace*, 2007.

[21]Forest, L., Hirsh, R., and Duda, K., "Results from the Crew Evaluation," Tech. rep., NASA Johnson Space Center, Houston, TX, June 2008.

[22]Prinzel, L., Kramer, L., Norman, R., Arthur, J., Williams, S., Shelton, K., and Bailey, R., "Synthetic and Enhanced Vision System for Altair Lunar Lander," *International Symposium on Aviation Psychology*, 2009.

[23]Forest, L., Cohanim, B., and Brady, T., "Human Interactive Landing Point Redesignation for Lunar Landing," *Proceedings of the IEEE Aerospace Conference*, Big Sky, MN, March 2008.

[24]Needham, J., *Human-Automation Interaction for Lunar Landing Aimpoint Redesignation*, Master's thesis, Massachusetts Institute of Technology, 2008.

[25]Chua, Z. and Major, L., "Task Modeling for Lunar Landing Redesignation," *AIAA InfoTech@Aerospace Conference*, 2009, In Press.

[26]Klein, G., *Sources of Power: How People Make Decisions*, MIT Press, MIT, 1998.

[27]Card, S., Moran, T., and Newell, A., *The Psychology of Human-Computer Interaction*, Lawrence Erlbaum, 1983.

[28]Olsen, J. and Olsen, G., "The Growth of Cognitive Modeling in Human-Computer Interaction Since GOMS," *Human-Computer Interaction*, Vol. 5, 1990, pp. 221–265.

[29]Tolbert, J. E., "Creation of a Mock Lunar Lander Environment," Tech. rep., Georgia Institute of Technology, Atlanta, GA 30332, 2009, For AE 4699: Undergraduate Research (K.M. Feigh).

[30]"Shuttle Set for Liftoff with 'Glass Cockpit'," Web news, April 21 2000, Accessed October 29, 2008.

[31]Machines., S., *FaceLAB4 User Operating Manual*, Seeing Machines, Canberra, Australia, 2006.

[32]"Take a Spin Around The Inside of Altair," February 2004.

[33]Schweiger, M., "Orbiter Simulator," Available online., September 2006.

[34]Homan, D., "EDGE (Engineering Doug Graphics for Exploration)," March 2007.

[35]Meyer, A., "X-Plane," Online., February 2009.

[36]Monsen, R., "EagleLander3D," October 2007.

[37]Wickens, C. and Carswell, C., "The proximity compatibility principle: Its psychological foundations and its relevance to display design," *Human Factors*, Vol. 37, 1995, pp. 473–494.

[38]Endsley, M. R., *Direct Measurement of Situation Awareness: Validity and Use of SAGAT*, chap. 7, Lawrence Erlbaum Associates, 2000, pp. 147–173.

[39]Cummings, M., Myers, K., and Scott, S. D., "Modified Cooper Harper Evaluation Tool for Unmanned Vehicle Displays," *UVS Canada Conference*, 2006.

[40]"Lunar and Planetary Institute (LPI)," Online., 2009, Accessed in January 2009.

[41]Marsh, A. K., "In-Flight Emergencies: Engine out," Available Online., January 1999.

[42]Center, M. S., "Apollo 12 Mission Report," Tech. rep., National Aeronautics and Space Administration, 1970.

Appendix

This section includes supplementary materials used in the experiment.

A. Scenario Mission Statements

Each scenario began with the sentence, "You have 30 seconds until Landing Point Redesignation (LPR)" and ended with "Our satellite imagery has captured this photograph of the landing area. You should expect the landing area to look similar. Results from the sensor scan will highlight dangerous hazards and provide terrain information". The following science missions were used:

- Scientists would like to return to either one of these landing sites. Results from the initial trip note the presence of an unusual mineral compound not normally found on Earth. They would like you to run some more extensive testing on-site and dig deeper into the regolith.
- Scientists believe large mineral deposits exist underneath the regolith in this landing site region. They would like you to run tests on-site and eventually bring back samples. These mineral deposits may be able to sustain future lunar bases.
- Scientists believe this is a mineral rich area, the result of the impact that left the crater nearby. They would like for you to land at this site (see arrow), run some on-site tests, and also help in the establishment of a lunar laboratory.
- Scientists want to examine the *ejecta* from nearby craters (see arrows). They believe the craters were formed at different times, but the *ejecta* will confirm the exact dates. Both sites will provide a wealth of knowledge. There is no preferred landing site.
- Scientists believe there are significant amounts of *dark mantle* deposits from pyroclastic eruptions. They would like you to take samples from this landing site (see arrow). These samples should confirm whether the minerals in these deposits can be harvested.
- Scientists believe there is a wealth of data about our universe at these two landing sites (see arrows). One is near a large crater and should contain information about a fallen meteorite. The other site is a *dark mantle* deposit from a pyroclastic eruption. Both sites are important. There is no preference between the two.
- Scientists believe this is an ideal site for setting up a lunar geology station (see arrow). However, before any foundations are laid, they would like you to test the soil for necessary mineral compositions such as iron, magnesium, and aluminium.
- Scientists have noted the presence of lunar objects such as *agglutinates* and believe this landing area (see arrow) should contain several high-quality samples. They would like you to retrieve these objects.
- Scientists are interested in learning more about this *straight rille*, sometimes formed by localized lava channels. They would like you to land at this site (see arrow) and run some on-site tests.
- Scientists want to investigate two sites of interest. The first potentially contains a *straight rille*, which are formed by localized lava channels. The second is an impact crater tha did not previously exist less than a decade ago. Both sites are of significant value. There is no preference between either site.
- Scientists would like to return to the place of a fallen meteorite. Astronauts previously erected a miniature experimental station (see arrow). This experimental station requires some repair and the collection of obtained soil samples.
- Scientists would like to study the chemical makeup of a meteorite that recently struck the Moon. Since landing directly in the crater poses significant problems, they would like for you to land at this specific location (see arrow). They believe this site is optimal for setting up a miniature experimental station.
- Scientists believe this crater field is home to vesicles, which are created from volcanic activity. They would like you to check whether these vesicles exist. If they do, you will also harvest the gas escaping from them and lay down the foundation for an experimental station.

- Scientists believe the regolith located in this crater field is particularly rich with information about our solar system, particularly, the Sun. They would like you to examine the isotopic composition of two lunar sites (see arrows). Both are scientifically valuable. There is no preference between landing sites.
- Scientists would like you to investigate this region to determine whether these craters are a *catena* (crater chain). Testing this hypothesis requires on-site measurements of the lunar soil chemical composition.
- Scientists would like you to set up a miniature experimental station at either one of these sites (see arrows). This station will analyze the lunar soil for mineral deposits and track the amount of sunlight at this location. The scientists will also need you to assist in equipment calibration.

B. Pre-briefing Questions

These questions were used to test the pilot's readiness for the LPR simulator.
Use Figure 35a to answer Questions 1-5. (Short answer)
Symbols on this figure: green ellipse, striped purple circles, boxed number, blue dot, yellow circles, numbered green circles, yellow tick marks, shaded red areas.
Use Figure 35b for Questions 6-8. (Multiple choice, short answer)
Use Figure 36 (behind this sheet) to answer Question 10.

1. Which symbol(s) describe area for which there is enough fuel?
2. Which symbol(s) indicate the point of interest?
3. How many landing options are available?
4. What do the purple striped circles represent?
5. What is the meaning of the shaded red areas?

6. What does the "20" signify?
7. What is the slope margin of the site in Figure 2? Use the scale of:
   - Safe - Far from Slope Threshold
   - Tolerable - Tolerable Slope Threshold
   - Warning - At Slope Threshold
8. What is the roughness margin of the site in Figure 35b? Use scale of:
   - Safe - Far from Roughness Threshold
   - Tolerable - Tolerable Roughness Threshold
   - Warning - At Roughness Threshold
9. Assume you have selected to sort the landing sites with respect to fuel efficiency. You push the FUEL button. What steps are now needed to select the second landing site as your final destination?
   - Step One:
   - Step Two:
10. Label the buttons along the bottom row using the following options: Button Names(Alphabetical Order): 1, 2, 3, A PR, Arm, BAL, Base, Fuel, POI, Safe

1.

2.

3.

4.

5.

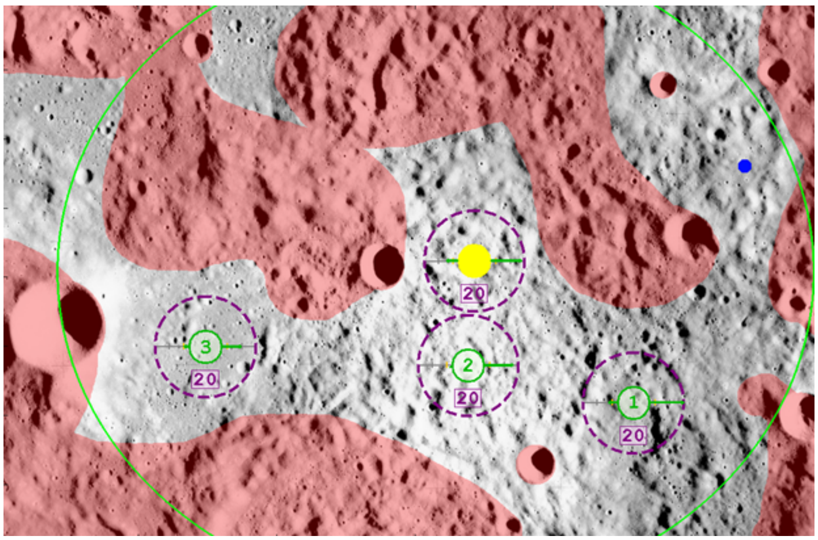
6.

7.

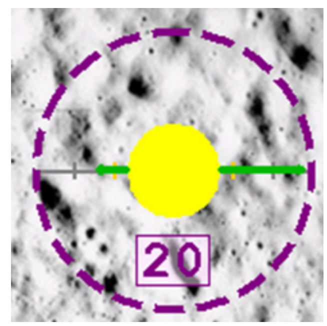
8.

9.

10.



(a) Use for Questions 1-5.



(b) Use for Questions 6-8.

Figure 35: Figures for Preliminary Questions

C. Debriefing Questions

Questions in quotations were asked of the pilot.

1. "Your recent run had [describe scenario]. These were the actions you took to land the vehicle [hand printed sheet for one of the runs]. Can you please explain why you selected this objective change?"
2. "Your recent run had [describe scenario]. These were the actions you took to land the vehicle [hand printed sheet for a second run]. Can you please explain why you selected this objective change?"
3. "Your recent run had [describe scenario]. These were the actions you took to land the vehicle [hand printed sheet for a third run]. Can you please explain why you selected this objective change?"
4. "If you were training another pilot to complete this task, what universal strategy would you suggest to them? Can you draw a flow chart?"
5. "Please list the decision attributes you were considering in selecting your final landing site." (examples: safety, specific rock formations, etc)."
6. Please draw a flow chart describing what strategy other pilots should use to complete the landing point redesignation task.
7. (Present Figure 37) "What types of information did you use to reach your decision? Please highlight and explain why."

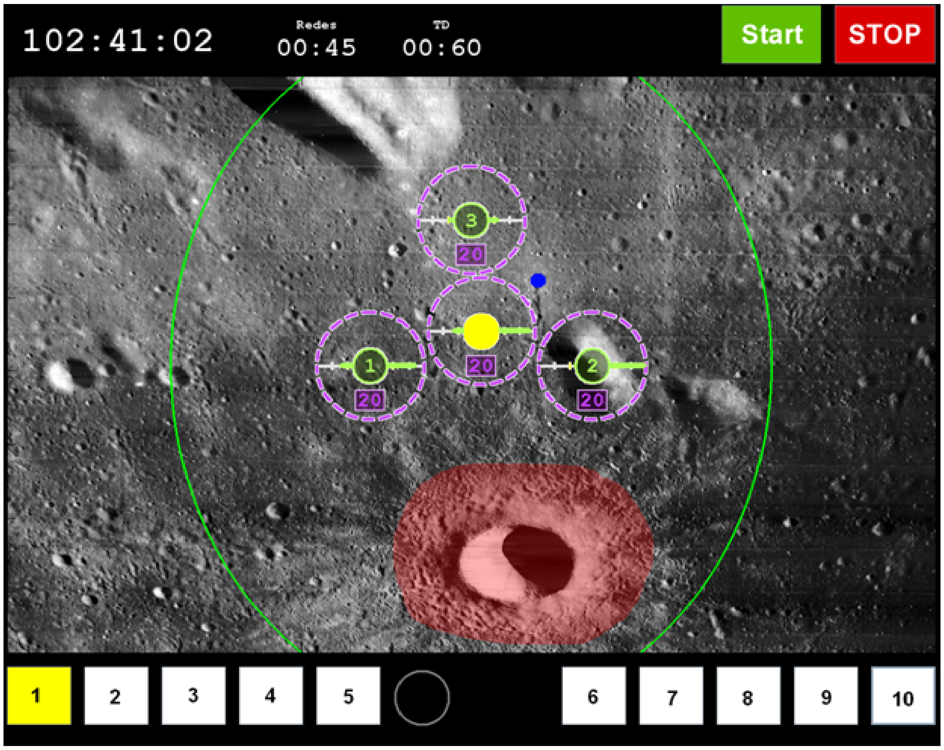


Figure 36: Use for Question 10.

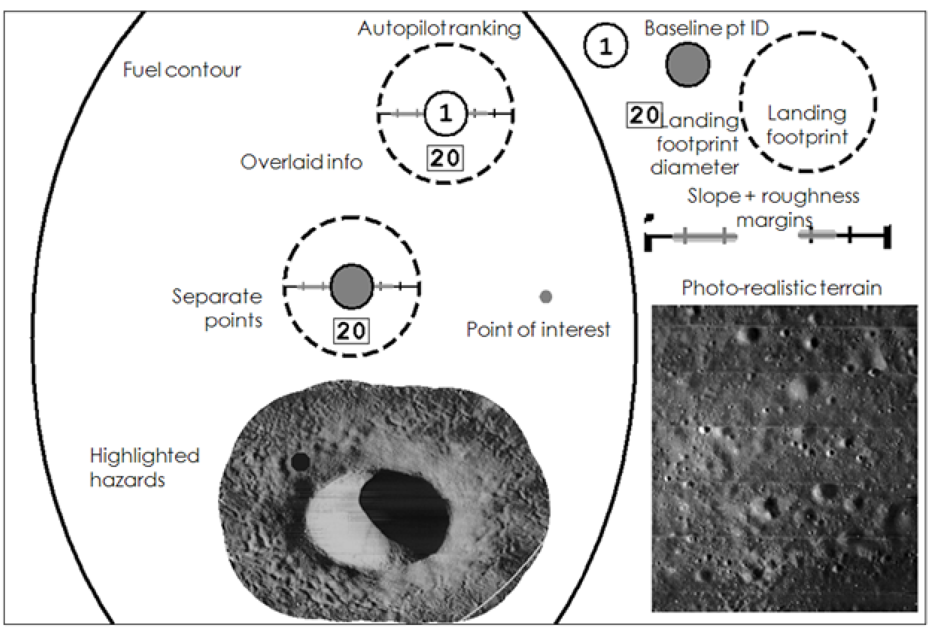


Figure 37: Types of symbols used.

8. “How useful was your window view?” (check all that apply)

    - **Not useful**: I believe the window distracted me from the task.
    - **Not useful**: I believe the window did not provide me the right information.
    - **Partly useful**: I glanced at the window once or twice, but relied mainly on the data presented to me.
    - **Partly useful**: I found the window view and the terrain data presented to be about equal in terms of useful information provided.
    - **Very useful**: I did not rely on terrain data, but only on the window to make my decision.
    - **Other**: please explain.

9. “Please rate the usefulness of the scale based on Figure 38.”

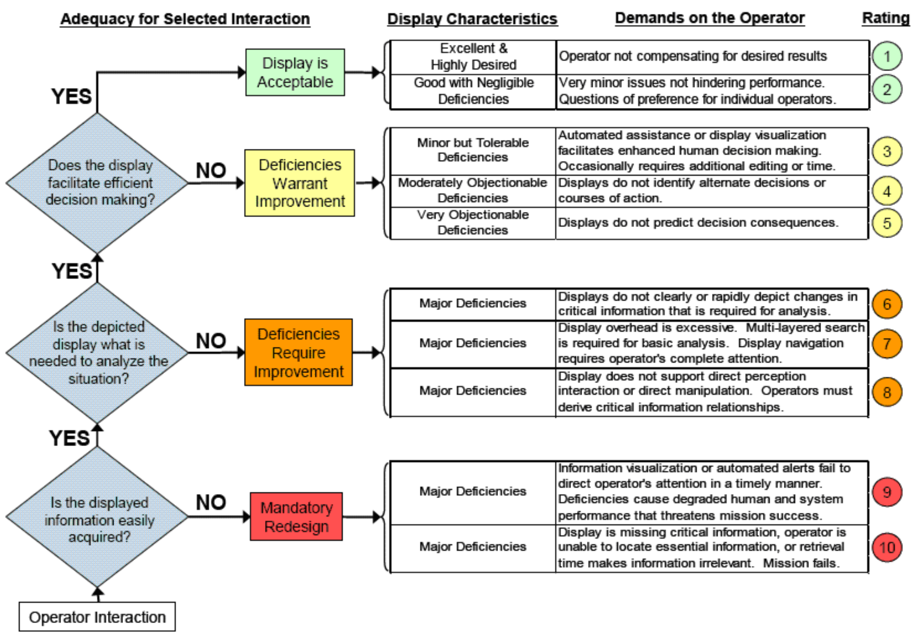


Figure 38: Modified Cooper-Harper Rating Scale.[39]

10. “What improvements would you suggest?”